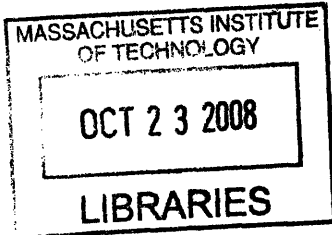Likelihood and Bayesian signal processing methods for the analysis of auditory neural and behavioral data

by

Anna A. Dreyer

M.Eng. Electrical Engineering and Computer Science
Massachusetts Institute of Technology, 2005

B.S. Electrical Engineering and Computer Science
Massachusetts Institute of Technology, 2003

Submitted to the Harvard-MIT Division of Health Sciences and Technology
in partial fulfillment of the requirements for the degree of
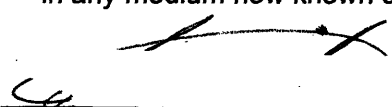
DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2008

Signature of Author: _____
Harvard-MIT Division of Health Sciences and Technology
August 1, 2008

Certified by: _____
Emery N. Brown, M.D., Ph.D.
Professor of Computational Neuroscience and Health Sciences and Technology
Massachusetts General Hospital Professor of Anaesthesia
Thesis Supervisor

Accepted by: _____
Martha L. Gray, Ph.D.
Edward Hood Taplin Professor of Medical and Electrical Engineering
Director, Harvard-MIT Division of Health Sciences and Technology

1

Likelihood and Bayesian signal processing methods for the analysis of auditory neural and behavioral data

by

Anna A. Dreyer

Submitted to the Harvard-MIT Division of Health Sciences and Technology on
August 1, 2008 in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Abstract

Developing a consensus on how to model neural and behavioral responses and to quantify important response properties is a challenging signal processing problem because models do not always adequately capture the data and different methods often yield different estimates of the same response property. The threshold, the first stimulus level for which a difference between baseline activity and stimulus-driven activity exists, is an example of such a response property for both neural and behavioral responses.

In the first and second sections of this work, we show how the state-space model framework can be used to represent neural and behavioral responses to auditory stimuli with a high degree of model goodness-of-fit. In the first section, we use likelihood methods to develop a state-space generalized linear model and estimate maximum likelihood parameters for neural data. In the second section, we develop the alternative Bayesian state-space model for behavioral data. Based on the estimated joint density, we then illustrate how important response properties, such as the neural and behavioral threshold, can be estimated, leading to lower threshold estimates than current methods by at least 2 dB. Our methods provide greater sensitivity, obviation of the hypothesis testing framework, and a more accurate description of the data.

Formulating appropriate models to describe neural data in response to natural sound stimulation is another problem that currently represents a challenge. In the third section of the thesis, we develop a generalized linear model for responses to natural sound stimuli and estimate maximum likelihood parameters. Our methodology has the advantage of describing neural responses as point processes, capturing aspects of the stimulus response such as past spiking history and estimating the contributions of the various response covariates, resulting in a high degree of model goodness-of-fit. Using our model parameter estimates, we illustrate that decoding of the natural sound stimulus in our model framework produces neural discrimination performance on par with behavioral data.

These findings have important implications for developing theoretically-sound and practical definitions of the neural response properties, for understanding information transmission within the auditory system and for design of auditory prostheses.

Thesis Supervisor: Emery Neal Brown
Title: Professor of Computational Neuroscience and Health Sciences and Technology
          Massachusetts General Hospital Professor of Anaesthesia

Acknowledgements

I am extremely grateful to the faculty, students and staff at MIT, Harvard University, Massachusetts Eye and Ear Infirmary and Massachusetts General Hospital who have made my graduate school experience enlightening, rigorous and enjoyable.

I would first like to thank my advisor, Dr. Emery Brown. Emery has served as an example to me in his knowledge and depth of understanding of statistical and mathematical concepts. However, his most unique quality, from my student perspective, is his enthusiastic nature and the joy with which he approaches science. His love for his work and encouragement is contagious, as evidenced by his students' deep respect and admiration for him. It has been such a pleasure for me to be mentored by him and to be a member of his laboratory. I would also like to express my gratitude to Dr. Bertrand Delgutte, the chair of my committee and my Master's thesis advisor. Bertrand taught me a deep appreciation for the experimental side of scientific research. The rigor with which he approaches problems and which he encourages in his students have certainly strengthened my abilities not just in science, but in challenging myself in all my pursuits. Dr. Uri Eden has been a wonderful addition to my committee. His deep understanding of the technical aspects of the problems detailed in this thesis have certainly improved the quality of the projects. Similarly, Dr. Patrick Wolfe has helped me to improve not only the technical, but the presentational aspects of my work. As a young scientist just beginning my career, I have also been greatly encouraged by his and Uri's examples as young, but extremely accomplished, faculty members. I would like to also thank my collaborators, David Anderson, Hubert Lim, David Anderson, Bryan Pfingst, Ann Graybiel, Tara Barnes Young, Yasuo Kabota, Ann Smith, Gabriela Czanner, and Kamal Sen who have either generously provided data for this thesis or have contributed to the algorithmic development.

My satisfaction with and enjoyment of my experience as a graduate student at the Harvard/MIT Department of Health Sciences and Technology were greatly affected by the wonderful interactions I have had with the faculty, staff and students in the program. In addition, I have enjoyed many helpful and enlightening scientific discussions with the members of the Neurostatistics laboratory. I particularly want to thank the following past and current members for their friendship and for serving as bouncing boards for ideas: Iahn Cajigas, Gabriela Czanner, Rob Haslinger, Antonio Molins, Michelle McCarthy, Dave Nguyen, and Sri Sarma. Thank you to Julie Scott for her organization of the Neurostatistics lab and for helping to maintaining its upbeat and friendly character.

Many of my most memorable and wonderful experiences as a graduate student were shared with my classmates in HST's Speech and Hearing Bioscience and Technology program. My classmates and I grew to be very close friends through our shared coursework and our social interaction outside of class. My memories as a graduate student will always be associated with them, particularly with Brad Buran, Amanda Graves, Erik Larsen and Jeannie Chung, Adrian KC Lee, Daryush Mehta and Andrea Gabert, Anton Peng, and Cara Stepp.

However, my greatest source of support has been my dearest, loving family: my parents, Alexander Dreyer and Elizabeth Chernin, and my husband, Richard Hanna. Their constant support of all of my endeavors, their encouragement, pride and love have sustained me and have enabled me to reach this tremendous milestone in my life. Several months before finishing my dissertation, our son Samuel was born, and he has added tremendous joy to our lives. I am so grateful for my family and I dedicate my accomplishments as a graduate student to them.

Table of Contents

Chapter 1

INTRODUCTION

Computational approaches to the analysis of auditory data generally employ two types of models. The first type of model is mechanistic in nature and describes the underlying mechanism mediating the neural or behavioral responses. Examples of such models in the auditory system include the model of the convergence of auditory nerve fibers on bushy cells in the ventral cochlear nucleus (Rothman et al., 1993) and models of coincidence detection cells found in the medial superior olive in mammals and the nucleus laminaris in birds (Dasika et al., 2006). The second type of model is empirical in nature and captures the statistical elements in the data rather than the mechanistic elements that may underlie the data. Examples include statistical models of neural response patterns in the cat lateral superior olive (Johnson et al., 1986; Turcott et al., 1994; Johnson, 1996) and models of behavioral performance based on auditory nerve discharge patterns (Siebert, 1968; Colburn, 1973, Heinz et al., 2001a, 2001b, 2001c). Empirical models are particularly suitable to describe responses to complex stimuli and to model responses in higher auditory centers where the underlying neural mechanisms are less well-understood (Popper and Fay, 1992; Wang et al., 1995; Theunissen et al., 2000; Depireux et al., 2001). These two classes of models are not completely detached and often contain some similar elements. For instance, findings from empirical models can provide the basis for the development of new mechanistic models, and mechanistic models can narrow the choices of data distributions in empirical modeling. In this thesis, we present several empirical models of auditory neural and behavioral data with influence from earlier mechanistic findings.

In presenting our empirical models of auditory data, our aim is to provide improved characterizations of response properties of the auditory system. To do so, we utilize likelihood and Bayesian principles, two general methods commonly used in the field of statistics for model estimation (Congdon, 2003) and the theory of inference. Likelihood methods are commonly used in statistical models to determine the quantities required to describe data and define the

7

amount of information, or, conversely, uncertainty, contained in data samples (Casella and Berger, 1990; Pawitan, 2001). The likelihood function describes how the data could have been generated from a probabilistic standpoint (Brown et al., 2003), and makes the statement that the hypothesis that assigned a greater probability to the observation will do a better job at predicting a specific event. Likelihood methods may be used in a variety of simple and very complex problems, ranging from linear regressions to nonlinear models with non-normal distributions. Bayesian methods, on the other hand, incorporate ideas about prior beliefs regarding the quantities that describe the data to augment the likelihood. Therefore, while likelihood methods are purely objective under the constraints of assumptions about the parameters of interest, Bayesian methods allow the inclusion of prior knowledge to augment the solution. Both methods provide principled and informative approaches for statistical modeling and facilitate the important task of inference from quantitative data (Pawitan, 2001; Congdon, 2003).

Parameter estimation in the likelihood framework is most commonly done by computing the maximum of the likelihood (ML) estimate of the parameter (Brown et al., 2003). The degree of the uncertainty present in the ML estimator is quantified using the Fisher Information, the Hessian of the log likelihood with respect to the parameter (Casella and Berger, 1990; Pawitan, 2001). ML estimates are consistent (converge in probability to the true value with a large sample size), asymptotically unbiased as sample size increases, efficient as sample size increases (have the smallest variance) and invariant (if $\hat{\theta}$ is ML of θ, then $f(\hat{\theta})$ is the ML of $f(\theta)$, where $f(\cdot)$ is a monotonic transformation) (Brown et al., 2003).

While analytical solutions for parameter estimation using Bayesian methods are possible, often, they are not computationally feasible. Techniques such as Markov Chain Monte Carlo (MCMC) methods instead provide a numerical solution by drawing samples from the distributions of interest. Sampling strategies, such as Gibbs sampling and Metropolis-Hastings algorithm, have been developed to provide efficient and thorough samples of the parameter

distributions (Congdon, 2003). Statistical estimates, such as the median and credible intervals, can then be computed from the sampled distributions.

We apply our likelihood-based framework to the analysis of neural data by modeling the responses of single-units as a point process, a set of discrete binary events that occur in continuous time. The point process model serves as a fitting description for neural data because action potentials are generally considered all-or-nothing for the purposes of cellular communication, in that each action potential is identical to the next (i.e. Rieke et al., 1999). More specifically, scientists consider differences in the increase of the membrane voltage not significant as long as the threshold potential is surpassed and the duration of the action potential as similar across multiple spikes (Kandel et al., 2000). Therefore, action potentials can be viewed as discrete events, because they are very short events in time, and are binary because a spike either occurred or did not occur, in that the cells membrane potential either exceeded or did not exceed the threshold voltage.

Recent research has showed how point process data can be analyzed within the likelihood framework (e.g. Brillinger, 1992; Brillinger, 1998; Barbieri et al., 2001; Kass and Ventura, 2001; Brown et al., 2003; Truccolo et al., 2005). The instantaneous firing probability of neurons can be described using a conditional intensity function, which provides a history-dependent extension of a non-homogeneous Poisson process (Brown et al., 2003). The conditional intensity function can be used to formulate the point process description in the likelihood framework and goodness-of-fit metrics appropriate for point process models can be applied.

In this thesis, we present two examples of auditory neural data analyzed within the likelihood point process framework (see Table 1). In Chapter 2, we model responses from cells in the primary auditory cortex to a brief current pulse to the inferior colliculus across multiple trials and multiple current levels. We model the dynamics from one level to the next using a state-space formulation (Kitagawa and Gersh, 1996), such that the dynamics modulating the

9

changes in responses from level to level are not observed, but influence the observable neural responses. We then compute the ML values of the parameters using the Expectation-Maximization algorithm, suitable for ML estimation of models with hidden internal dynamics (Dempster et al., 1977; Tanner, 1991; McLachlan and Krishnan, 1997). Using our estimated model parameters, we approach the problem of threshold estimation from our likelihood framework, by finding the current level for which the spontaneous neural activity is just significantly different from stimulus-induced activity. We compare our dynamic point process description of neural firing to a static description based on a homogeneous Poisson process and show that the Poisson description is less appropriate in describing the properties of the data. In addition, we illustrate that the threshold inferences using our framework are significantly lower than those using current approaches, a finding which has strong implications for understanding auditory central processing and improving the design of auditory prostheses.

We present another example of a model class of auditory data in the likelihood-based point process framework by describing responses in the auditory nerve, cochlear nucleus and field L, the avian analogue of the primary auditory cortex, to natural stimuli in Chapter 4 (see Table 1). Understanding what aspects of the natural stimulus make the neuron spike has been a challenging problem because of the complexity of the stimuli and particularly in higher areas of the auditory system, where cells receive inputs from many nuclei with different axonal and filtering delays (Aitkin, 1990; Webster et al., 1992; Ehret and Romand, 1996). We show that the problem of finding an appropriate stimulus description can be approached in our framework by using the empirical mode decomposition (Huang et al., 1998) to describe the stimulus, a technique particularly appropriate for nonlinear and non-stationary stimuli. The ML values of the model parameter estimates are computed using the generalized linear model (GLM) framework to capture the dynamic firing patterns of the cells in response to natural stimuli (i.e. Truccolo et al., 2005). We illustrate a high degree of goodness-of-fit for our models and show that our likelihood-based framework leads to neural discrimination on par with behavioral data.

10

	CHAPTER II	CHAPTER III	CHAPTER IV
Experiment	Neural spike recordings in primary auditory cortex in response to electrical stimulation of the ipsilateral inferior colliculus	Behavioral responses in response to electrical stimulation in the inferior colliculus	Neural spike recordings in the auditory nerve, cochlear nucleus and field L in response to natural acoustic stimuli
Data	Multiple trials of spike trains at multiple current levels	Binary responses and reaction times at multiple levels	Multiple trials of spike trains at 1 level
Model	Point process observation model with hidden state-space dynamics	Binomial and Gaussian or lognormal observation model with hidden state-space dynamics	Point process observation model
Estimation	Maximum Likelihood estimation using Expectation-Maximization algorithm	Bayesian estimation using MCMC methodology in BUGS	Maximum Likelihood estimation of Generalized Linear Model using Iteratively Reweighted Least Squares

Table 1: Summary of experiments, data, models and estimation procedures in thesis. Chapters 2 and 4 use likelihood-based methods, while chapter 3 uses Bayesian methods for parameter estimation and inference.

Although likelihood approaches could have also been used, we chose to instead present

a Bayesian approach for formulating a dynamic model of auditory behavioral responses in

Chapter 3 (see Table 1). We select the Bayesian framework and compute parameter estimates

using MCMC techniques in order to test several models without having to augment the

estimation algorithm with each new model (Congdon, 2003; Smith and Brown, 2007). In

Chapter 3, we analyze psychometric functions, the relationship between stimulation level and a

proxy for the internal representation of the stimulus intensity, such as percent correct detection.

Similarly to Chapter 2, we model the dynamics from one level to the next using a state-space

formulation, and illustrate the process of threshold inference, by estimating the level at which a

specific level of detection is reached, as well as more probabilistic definition as the first level for

which there is reasonable certainty that performance exceeds baseline. In addition, we illustrate

how the Bayesian framework allows for incorporation of other metrics of the internal

representation of the stimulus intensity, such as reaction time, and for model parameters using

multiple proxies to be estimated simultaneously. Lastly, we illustrate how to extend this model to scenarios where multiple subjects within multiple population groups are studied.

A very important aspect of our methodology, both using the likelihood and Bayesian framework is the ability to estimate the joint distribution of the data. This methodology allows us to bypass the hypothesis testing framework completely when making inferences. Rather than testing the null hypothesis under two marginal distributions of the parameter of interest, our methodology allows us to construct confidence bounds for our parameter estimates. We can then use these confidence bounds to compare parameter estimates in various scenarios. By bypassing the hypothesis testing methodology, we also avoid multiple comparison tests that the hypothesis testing methodology demands, but that are often neglected.

Dynamic likelihood and Bayesian estimation models of auditory neural and behavioral data have important implications for understanding the dynamics of the auditory system and for design of auditory prosthesis as aids for the deaf and hard-of-hearing. We summarize the thesis results and implications in Chapter 5.

REFERENCES

Aitkin L (1990) *The auditory cortex: structural and functional bases of auditory perception.* New York: Chapman and Hall.

Barbieri R, Quick MC, Frank LM, Wilson MA, Brown En. Construction and analysis of non-Poisson stimulus response models of neural spike train activity. *J Neurosci Methods* 105: 25-37.

Birnbaum A (1962) On the foundations of statistical inference. *J Am Stat Assoc* 57: 269-306.

Brillinger DR (1992) Nerve cell spike train data analysis: a progression of techniques. *J Amer Stat Assoc* 87: 260-271.

Brillinger DR (1988) Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol Cyber* 59: 189-200.

Brown EN, Barbieri R, Eden UT, Frank LM (2003) Likelihood methods for neural data analysis. In: Feng J, ed. Computational Neuroscience: A Comprehensive Approach. London: CRC, Chapter 9: 253-286.

Casella G, Berger RL (1990) *Statistical inference.* Belmont, CA: Duxbury.

Colburn HS (1981) Intensity perception: Relations of intensity discrimination to auditory-nerve firing patterns. Internal Memorandum, Res. Lab. Electron. MIT, Cambridge, MA.

Congdon P (2003) *Applied Bayesian Modelling.* John Wiley and Sons Ltd., Chichester, United Kingdom.

Dasika VK, White JA, Colburn HS (2006) Simple models show the general advantage of dendrites in coincidence detection. *J Neurophysiol* 97: 3449-3459.

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J Roy Statist Soc B* 39: 1-38.

Depireux DA, Simon JZ, Klein DJ, Shamma SA (2001) Spectro-temporal receptive field characterization with dynamic ripples in ferret primary auditory cortex. *J Neurophysiol* 85: 1220-1234.

Ehret G, Romand R (Eds) (1996) *The central auditory system.* New York: Oxford University Press.

Heinz MG, Colburn HS, Carney LH (2001a) Evaluating Auditory Performance Limits: I. One-Parameter Discrimination Using a Computational Model for the Auditory Nerve. *Neural Comput* 13: 2273–2316.

Heinz MG, Colburn HS, Carney LH (2001b) Evaluating Auditory Performance Limits: II. One-Parameter Discrimination with Random-Level Variation. *Neural Comput* 13: 2317–2338.

13

Heinz MG, Colburn HS, Carney LH (2001c) Rate and timing cues associated with the cochlear amplifier: Level discrimination based on monaural cross-frequency coincidence detection. *J Acoust Soc Am* 110: 2065-2084

Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Lio HH (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc Math Phys Eng Sci* 454: 903-995.

Johnson DH, Tsuchitani C, Linebarger, DA, Johnson MJ (1986) Application of a point process model to responses of cat lateral superior olive units to ipsilateral tones. *Hear Res* 21: 135-159.

Johnson DH (1996) Point process models of single-neuron discharges. *J Comp Neurosci* 3: 275-299.

Kandel ER, Schwartz JH, Jessell TM (2000) *Principles of Neural Science*. New York: McGraw-Hill Publishing Co.

Kass RE, Ventura VA (2001) A spike train probability model. *Neural Comput* 13: 1713-1720.

Lee Y, Nelder JA, Pawitan Y (2006). Generalized linear models with random effects: unified analysis via H-likelihood. Boca Raton: Chapman & Hall/CRC.

McLachlan GJ, Krishnan T (1997) The *EM Algorithm and Extensions*. New York: John Wiley & Sons.

Pawitan Y (2001) *In All Likelihood: Statistical Modeling and Inference Using Likelihood*. New York: Oxford Univ. Press.

Popper AN, Fay RR (Eds) (1992) *The Mammalian auditory pathway: neurophysiology*. New York: Springer.

Rieke F, Warland D, van Steveninck RDR, Bialek W (1999) *Spikes: Exploring the neural code* Cambridge, MA: The MIT Press.

Rothman JS, Young ED, Manis PB (1993) Convergence of auditory nerve fibers onto bushy cells in the ventral cochlear nucleus: implications of a computational model. *J Neurophysiol* 70: 2562-2583.

Smith AC, Wirth S, Suzuki W, Brown EN (2007) Bayesian analysis of interleaved learning and response bias in behavioral experiments. *J Neurophys* 97: 2516-2524.

Tanner MA (1991) *Tools for statistic inference*. New York: Springer-Verlag.

Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci* 20: 2315-2331.

Truccolo, W. Eden, U.T., Fellows, M.R., Donoghue, J.P. and Brown, E.N (2005) A point process framework for relating neural spiking activity to spiking history, neural ensemble and extrinsic covariate effects. *J Neurophysiol* 93: 1074-1089.

Turcott RG, Lowen SB, Li E, Johnson DH, Tsuchitani C, Teich MC (1994). A nonstrationary Poisson point process describes the sequence of action potentials over long time scales in the lateral-superior-olive auditory neurons. *Biol Cybern* 70: 209-217.

Webster DB, Popper AN, Fay RR (Eds) (1992) *The Mammalian auditory pathway: neuroanatomy.* New York: Springer.

Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J Neurophysiol* 74: 2685-2706.

Chapter 2

Point process statistical framework for auditory threshold estimation

1.   ABSTRACT

The neural threshold, defined as the stimulus level for which the spike activity initially differs from the spontaneous response, is an example of an inference for which definitions and methodologies vary. Most current methods do not consider the point process nature of the neural spiking activity, the relation between neural responses at different current levels, use the hypothesis testing framework while not correcting for the multiple comparisons required to determine the detection threshold. We developed a point process state-space generalized linear model statistical framework that considers both dependence of neural activity on spiking history and the dynamics of the responses across all stimulus levels to estimate neural thresholds. We applied our method to 30 multi-unit recordings in primary auditory cortex (A1) in response to electrical stimulation of the inferior colliculus across 19 current levels. We fit the model of neural responses to current stimulation by maximum likelihood using an EM algorithm. Our flexible framework allowed for threshold inferences based on single trials between two different levels, multiple trials at two different levels (the stimulus level versus the baseline level), multiple trials at multiple levels, and single trials from multiple neurons between two different levels. In comparison to several current methods, our methods improved threshold detection by at least 2 dB. Our findings have important implications for theoretically sound and practical definitions of neural threshold, for other statistical inferences based on neural data.

2. INTRODUCTION

Accurate estimation of neural response properties, such as neural threshold, is important for understanding neural circuits and classifying neural responses. The neural threshold has been used to compare response properties between different nuclei (Warchol and Dallos, 1990), across animals with genetic mutations (Taberner and Liberman, 2005; Oliver et al., 2006) and with electrical versus acoustic stimulation (Greenberg, 1998; Shepherd and Javel, 1997; Shepherd et al., 1993). Auditory prostheses also require accurate and sensitive neural response estimates to determine safe, but also effective current levels for activation (Lim and Anderson, 2007).

To date, estimation techniques for these neural response properties vary widely, making comparisons across species and studies difficult. For example, commonly used methods for threshold estimation include the visual method (Bietel et al 2002; Ramachandran et al., 1999; Syka et al. 2000; Vollmer et al., 2001), the standard deviation or number of spikes above spontaneous activity method (Raggio and Schreiner, 1994; Moxon, 1974, Witte et al., 1998; Sumner et al., 2005), neural networks (Middlebrooks, 2004), the rate-level function (RLF) method (Bierer and Middlebrooks, 2002) and methods using signal detection theory (STD) (Green and Swets, 1966; Britten et al., 1992; Lim and Anderson, 2006). Most of these methods do not consider the point process nature of the spiking activity, the relation between neural responses at different current levels, and formulate the problem in the hypothesis testing framework while not correcting for the multiple hypothesis tests required to determine the detection threshold. Although at least one study has estimated threshold based on the degree of temporal precision of spiking (Lane and Delgutte, 2005), most methods estimate the threshold based on the average firing rate of the neuron. Since many neurophysiological studies estimate not only the neural threshold, but also different features of neural coding relative to threshold, the lack of a consistent threshold method can lead to different functional characterizations of a neural system.

The focus of this paper is to present a more unified method for computing neural thresholds. We developed a state-space generalized linear model (SS-GLM) that captures the rate-based and temporal dynamics of the neural response across several levels of stimulation and takes into account the neural spiking history based on the point process nature of spikes. Our approach follows that of the SS-GLM model for capturing within-trial dynamics recently reported in Czanner et al. (2008). The model parameters are estimated by maximum likelihood (ML) using the expectation maximization (EM) algorithm. We use as goodness-of-fit metrics (Kolmogorov-Smirnov (KS) plots and autocorrelation function applied to rescaled intervals (Brown et al., 2002) designed for point-process data to assess degree of agreement between the model and the data. Once our encoding models are estimated with adequate model fit, we use decoding techniques to infer neural discrimination based on single trials from two levels, many trials at two levels and many trials at many levels. However, our model framework, which fully characterizes the neural responses, can be extended to evaluate other coding properties represented by the spiking activity. We discuss some of these features as well as how our methodology can be used to optimize the design of experiments to potentially reduce the number of trials and the length of the recording time.

3.   METHODS

In the following section, we detail our methods for making neural threshold inferences. Because well-fit models are needed characterize neural dynamics accurately, we first describe our SS-GLM model framework and our goodness-of-fit metrics. Next, using the fitted models, we explain how thresholds are estimated by computing estimates based on single trials from two levels, multiple trials from two levels and many trials from many levels, as well as combining responses across multiple neurons.

3.1   State-space generalized linear model of neural responses at different levels of stimulation

We assume that neural data can be modeled as a point process, a time-series of binary events occurring in continuous time, where the binary events represent a spike (1) or no spike (0) (Brown et al, 2003; Daley and Vere-Jones, 2003; Brown, 2005). On a recording interval of $(0, T]$, the counting process $N(t)$ represents the cumulative spike count on the interval $(0, t]$ at times $0 < u_1 < u_2 < u_3 < \cdots < u_J \leq t \leq T$. Consequently, a model of the point process can be completely characterized by the conditional intensity function (CIF) that defines the spike rate at every point in time as

$$\lambda(t|H(t)) = \lim_{\Delta \to 0} \frac{\Pr\left[N(t + \Delta) - N(t) = 1|H(t)\right]}{\Delta} \tag{2.1}$$

where $H(t)$ represents the spiking history until time t (Brown et al., 2003; Daley and Vere-Jones, 2003; Brown, 2005). Equation (2.1) generalizes the definition of the rate function of a point process to one that is history dependent (Daley and Vere-Jones, 2003).

We formulate a specific model of the CIF that reflects common paradigms for recording spike trains for threshold estimation in auditory experiments. We assume that we collect neural spike trains for $1, \dots, K$ trials and $1, \dots, V$ stimulus levels. We divide the recording interval $(0, T]$ into B time steps, such that the length of each time interval is $\Delta = TB^{-1}$ and each time step is

indexed as $b\Delta$. In this case, the discrete version of the CIF at time $b\Delta$ is $\lambda_{v,k}(\Delta b | H_{v,k,b})$ where $H_{v,k,b\Delta}$ represents the spiking history at stimulus level v and trial k up to a time bin b We select a sufficiently large B so that $\lambda_{v,k}(\Delta b | H_{v,k,b})\Delta$ approaches the probability of seeing one spike in the time interval Δ We can also represent the parametric form of the conditional intensity function at time $b\Delta$ as $\lambda_{v,k}(\Delta b | \Psi, H_{v,k,b})$ where Ψ represents parameters to be estimated. We represent the spike times as a series of unit pulse functions $n_{v,k,b}$ that takes on the value of 1 if a spike occurs in the interval $((b-1)\Delta, b\Delta]$ for stimulus level v and trial k, and takes on a value of 0 otherwise. We represent the binned spike times at stimulus level v and trial k as $N_{v,k} = \{n_{v,k,1}, \dots, n_{v,k,B}\}$ and represent all the data at a given stimulus level v as $N_v = \{N_{v,1}, \dots, N_{v,k}\}$ and the entire dataset as $N_{1:V} = \{N_1, \dots, N_V\}$.

We use likelihood methods for fitting the CIF to the data. To use these methods, one must derive the joint probability density for the spike train. The joint density, or likelihood, will be used in defining the encoding model but also in decoding the threshold inference. If we select a sufficiently small Δ such that the probability of seeing more than one spike in this binwidth approaches zero, the joint probability can be written as the product of Bernoulli independent events (Truccolo, et al., 2005):

$$
\begin{aligned}
\mathrm{L}[\Psi | N_{1:V}] = \mathrm{Pr}\,[N_{1:V} | \Psi] = \\
\prod_{v=1}^{V} \prod_{k=1}^{K} \prod_{b=1}^{B} \frac{[\lambda_{v,k}(b\Delta | H_{v,k,b}, \Psi)\Delta]^{n_{v,k,b}}}{[1 - \lambda_{v,k}(b\Delta | H_{v,k,b}, \Psi)\Delta]^{1-n_{v,k,b}}} \\
+ o(\Delta)
\end{aligned}
\tag{2.2}
$$

where $o(\Delta)$ represents the probability of seeing two or more spikes on the interval $\big((b-1)\Delta, b\Delta\big]$ and Ψ represents the parameters to be estimated. If we select Δ sufficiently small (for example, the refractory period of a neuron) such that the occurrence of two or more spikes is highly unlikely, then this probability tends to zero:

$$
\lim_{\Delta \to 0} \frac{o(\Delta)}{\Delta} \to 0
\tag{2.3}
$$

Making the above assumption, one can show that the probability mass function or likelihood has the form of the likelihood for a generalized linear model (GLM) under a Poisson probability model and a log link function (Truccolo, 2005; Czanner et al., 2008).

The neural threshold represents an identifiable response to a stimulus, and we therefore assume that the stimulus modulates the firing behavior of the neuron. However, as suggested by Dayan and Abbot (2001), an analysis of neural responses should also simultaneously characterize the neuron's biophysical properties. We model the biophysical properties as the dependence of the neural response on its own spiking history, $H_{v,k,b}$, and assume that this dependence is the same for all stimulus levels. This assumption leads to the following parameterization of the CIF:

$$\lambda_{v,k}\left(b\Delta|\theta_v, H_{v,k,b}, \gamma\right) = \lambda_{v,k}^S(b\Delta|\theta_v)\lambda_{v,k}^H\left(b\Delta|H_{v,k,b}, \gamma\right). \tag{2.4}$$

where $\lambda_{v,k}^S(b\Delta|\theta_v)$ is the stimulus effect and reflects the component of the CIF representing sensitivity to the stimulus (with units of spikes per second) and parameterized by θ_v, and the second component, $\lambda_{v,k}^H\left(b\Delta|H_{v,k,b}, \gamma\right)$, models the spiking history effect as a unit-less scaling coefficient by convention (Kass and Ventura, 2001; Truccolo, 2005) and parameterized by γ, which is independent of stimulus level. We select this parametric form of the model following several studies that have isolated the stimulus portion from the history portion of neural responses in different neural centers (Gaumond et al., 1982; Johnson and Swami, 1983; Miller, 1985; Johnson et al., 1986; Miller and Mark, 1992; Johnson, 1996; Brown et al., 2002; Brown et al, 2001; Truccolo et al., 2005; Czanner et al., 2008).

To model the stimulus effect on the neural spiking activity, we assume that the first component of the CIF has the form:

$$\lambda_{v,k}^S(b\Delta|\theta_v) = \exp\left(\sum_{r=1}^{R}\theta_{r,v}g_r(b\Delta)\right) \tag{2.5}$$

where $g_r(b\Delta)$ are general basis functions and R represents the dimensionality of the estimated

parameter $\theta_v = \{\theta_{v,1}, \ldots, \theta_{v,r}, \ldots, \theta_{v,R}\}$. For this dataset, we assume that the trial variations in the timing of the spikes can be captured in the history component of the response and we therefore assume a trial-independent parameterization of the stimulus parameter θ_v. Although the form of the $g_r(b\Delta)$ basis functions can be quite general, in this application we assume $g_r(b\Delta)$ to be unit pulse functions of equal length, such that:

$$g_r(b\Delta) = \begin{cases} 1 \; if \; b = \dfrac{(r-1)B}{R}, \ldots, \dfrac{rB}{R} \\ 0 \quad otherwise \end{cases}. \qquad (2.6)$$

When the unit pulse function is selected for each basis $g_r(b\Delta)$, the stimulus-effect represents a level-dependent peri-stimulus time histogram (PSTH). In fact, it can be shown that the ML estimate of the CIF with no spiking history effect would lead to the PSTH (Czanner et al., 2008). For each stimulus level $v = 1, \ldots, V$, θ_v can potentially take on different values and the state-space model defines the relationship among each θ_v across levels. For other applications, other choices of $g_r(b\Delta)$ are possible, including spline functions and polynomials; however, we selected unit pulse functions to draw the analogy to a level-dependent PSTH for a pulse input.

To model the effect of the spiking history on the current neural spiking activity, we assume the following form for the second component of the CIF:

$$\lambda_{v,k}^H(b\Delta | H_{v,k,b}, \gamma) = \exp\left(\sum_{j=1}^{J} \gamma_j n_{v,k,b-j} \right) \qquad (2.7)$$

where γ_j relates the current spiking activity to spiking activity during the previous $j = 1, \ldots, J$ intervals. We capture the history dependence $H_{v,k,b}$ at a resolution of Δ (typically 1 ms consistent with the refractory period of a neuron). Here, we assume that the spiking history effect reflects the intrinsic dynamics of the cell that do not change when the stimulus is applied. The dependence of spiking behavior on past spiking history can capture observed phenomena such as self-inhibiting behavior (refractory period), self-exciting behavior and bursting (Truccolo et al., 2005). Although we assume that the history parameters do not change with stimulus

22

level, and hence is not used to calculate neural thresholds, the modeling the history is still important because it prevents left-out parameter bias related to bursting and self-exciting behavior.

Therefore, combining equations (2.4) through (2.7) , we obtain the following complete form of the CIF for our problem:

$$
\begin{aligned}
&\lambda_{v,k}\big(b\Delta\big|\theta_v,\gamma,H_{v,k,b}\big)\\
&= \exp\left(\sum_{r=1}^{R}\theta_{v,r}g_r(b\Delta)\right)\exp\left(\sum_{j=1}^{J}\gamma_j n_{v,k,b-j}\right)
\end{aligned}
\tag{2.8}
$$

for trials $k = 1, ..., K$ and levels $v = 1, ..., V$, which defines our likelihood function described by a GLM (Brillinger, 1988; McCullagh and Nelder, 1989; Kass and Ventura, 2001; Truccolo et al. 2005; Czanner et al., 2008). Modeling the CIF as a product of exponentials ensures that the firing rate is always strictly positive.

We complete the definition of the state-space model (Kitagawa and Gersh, 1996) by characterizing the relationship between spiking activity across stimulus levels $v = 1, ..., V$. Empirical observations suggest that with fine level steps, the response to a subsequent stimulus level closely resembles the response to a previous stimulus level. Therefore, we assume that the parameters $\theta_{v:V,r:R} = [\theta_{1,r:R}, ..., \theta_{V,r:R}]$ are governed by a random walk at each time bin:

$$
\theta_{v,r} = \theta_{v-1,r} + \epsilon_v
\tag{2.9}
$$

for levels $v = 1, ..., V$ where $\epsilon_v = [\epsilon_{v,1}, ..., \epsilon_{v,R}]$ is an R-dimensional Gaussian random variable with 0 mean and an unknown covariance matrix Σ. In addition, we assume that the initial values of the $\theta_{v,r}$ parameters, θ_0, are unknown. Equations 2.4 through 2.9 constitute our state-space generalized linear model. This model can be used to characterize the neural dynamics and to do inference on neural thresholds, provided that we can estimate the unknown parameters, $\Psi = (\theta_0, \Sigma, \gamma)$. Because the values of the random walk, $\theta_{v,r}$, are unobservable, we use the EM

algorithm to compute the parameter estimates by maximum likelihood (Dempster et al., 1977; Tanner, 1991; McLachlan and Krishnan, 1997). This technique has been applied to solving behavioral (Smith and Brown, 2003; Smith et al. 2004; Smith et al., 2007) and neurophysiological (Czanner et al., 2008) problems. The EM algorithm is an iterative procedure consisting of two steps at each iteration: an expectation step and a maximization step. In the expectation step, the algorithm estimates the expected log-likelihood given a current estimate of the unknown parameters, where the expectation is computed with respect to the hidden observations. In the maximization step, the parameter estimates are updated by maximizing the expected-log likelihood. The joint posterior distribution of the hidden processes is approximately Gaussian, whose moments are calculated as a byproduct in the last expectation step of the EM algorithm. The EM estimates of the unknown parameters have approximately Gaussian distributions as well (Dempster et al., 1977; Tanner, 1991). We calculate their asymptotic covariance matrix as the inverse of the expected information matrix (McLachlan and Krishnan, 1997; Czanner et al., 2008). See Appendix A for implementation of EM algorithm for our problem.

3.2. *Goodness-of-fit metrics*

Before making an inference from a statistical model, it is crucial to assess the extent to which the model describes the data. Standard metrics of the quantitative agreement between a proposed continuous valued model and continuous data, such as sum of squared errors, cannot be applied to point process data such as spike trains. An alternative method, utilizing the time-rescaling theorem (Ogata, 1988; Brown et al., 2001; Brown et al., 2002) has been specifically designed to conduct goodness-of-fit assessments on point process data. The idea is to transform the point process measurements into a continuous measure appropriate for goodness-of-fit assessment.

Once a parametric form of the CIF has been estimated, the rescaled interspike interval times can be computed as:

$$z_{v,k,m} = 1 - \exp\left(- \int_{u_{v,k,m-1}}^{u_{v,k,m}} \lambda_{v,k}(u|\widehat{\Psi})du\right) \qquad (2.10)$$

where $u_{v,k,m}$ is the spike time of the spike m at stimulus level v and trial k for $m = 1, \dots, M_{v,k}$ and

$M_{v,k}$ is the total number of spikes in stimulus level v and trial k. We assume that $u_{v,k,0} = 0$. We

represent the ML estimates from the modeled CIF as $\widehat{\Psi} = (\widehat{\theta_0}, \hat{\gamma}, \widehat{\Sigma})$. The time-rescaling states

that if the estimated CIF approximates the true CIF well, the rescaled intervals $z_{v,k,m}$ will be

independent and uniformly distributed on the interval [0,1]. Therefore, a comparison between a

random sample from a uniform distribution and the rescaled intervals can assess how well the

model corresponds to the true CIF. To assess the uniformity of the transformation, the rescaled

intervals $z_{v,k,m}$ can be ordered from smallest to largest $(z^*_{v,k,m})$ and plotted against the values of

a uniform cumulative density distribution defined as $c_{m^*_{v,k}} = (m^*_{v,k} - \frac{1}{2})/M^*_{v,k}$ for spike times

$m^*_{v,k} = 1, \dots, M^*_{v,k}$. This is known as a Kolmogorov-Smirnov (KS) plot (Johnson & Kotz, 1970;

Brown et al, 2002). If the model is correct, the ordered rescaled times will lie along a 45-degree

line. Confidence bounds can be calculated according to the KS statistic, and for medium to

large data sizes, 95% confidence bounds can be well approximated by $c_{m^*_{v,k}} \pm \frac{1.36}{\sqrt{M^*_{v,k}}}$ (Johnson

and Kotz, 1970).

To assess independence of the rescaled intervals, the uniform-rescaled intervals can be

transformed to a Gaussian distribution as $z^G_{v,k,m} = \phi^{-1}(z_{v,k,m})$, where $\phi(\bullet)$ is a cumulative

distribution function of a Gaussian random variable with zero mean and unit variance. If the

Gaussian rescaled times $z^G_{v,k,m}$ are uncorrelated up to a certain lag, then they are also

independent up to that lag, since lack of correlation implies independence for a Gaussian

random variable. We assess the correlation of the rescaled times $z^G_{v,k,m}$ by plotting the

autocorrelation function and its confidence intervals (Box et al., 1994). If the autocorrelation

function remains generally within the confidence bounds up to a certain lag, then we can be

confident that the original times are also independent up to that lag of consecutive interspike intervals. The confidence bounds are calculated using Monte Carlo procedures with simulated draws from identically distributed standard Gaussian random variables.

3.3. Estimation of stimulus effect on the spiking activity

Estimation of the stimulus effect on the spiking activity is crucial to threshold estimation because the stimulus effect is the portion of the neural response that changes with stimulus level and with time within a trial. We estimate the stimulus effect based on the invariance principle of ML. The invariance property implies that if $\widehat{\Psi} = (\hat{\theta}_0, \hat{\gamma}, \hat{\Sigma})$ is the ML estimate of Ψ, then any function f of the ML estimator $f(\widehat{\Psi}) = f(\hat{\theta}_0, \hat{\gamma}, \hat{\Sigma})$ is also the ML estimate of $f(\Psi)$ (Pawitan, 2001). All the optimality properties of the ML estimation method that hold for $\widehat{\Psi}$ also hold for $f(\widehat{\Psi})$. Therefore, once the parameter ML estimates are determined, we can combine the parameter estimates to arrive at the ML estimate of the CIF, or any component thereof. For instance, The ML estimate of the stimulus effect $\lambda_{v,k}^S(b\Delta|\theta_v)$ at time $b\Delta$ and stimulus level v trial k is:

$$\lambda_{v,k|V}^S\left(b\Delta|\theta_{v|V}\right) = \exp\left(\sum_{r=1}^{R} \theta_{r,v|V} g_r(b\Delta)\right) \qquad (2.11)$$

where $\theta_{v|V}$ is the estimate of the hidden state, θ_v, based on all K trials and on the ML and the ML estimates of $\widehat{\Psi}$. Details of the algorithm and methodology for obtaining simultaneous confidence bounds for a similar problem are detailed in Appendix B.

3.4. Threshold inference using multiple trials at multiple stimulus levels

After estimating the model parameters and ascertaining goodness-of-fit, the principal objective is to make threshold inferences from the model. The simplest, non-model-based, way of determining threshold is to look at the dot raster responses across multiple trials and levels and visually detect the stimulus level at which the response just begins to differ from baseline. This visual method of threshold detection uses data from multiple trials and levels but allows for

26

subjectivity in the interpretation. The multiple-trial multiple-level (MTML) threshold inference also uses information across many trials and levels, but without subjectivity. The best know example of this class of MTML estimates without subjectivity are the rate-level function method, which sums across trial and time to arrive at one mean firing rate as a function of stimulus level (Bierer and Middlebrooks, 2002) and the standard deviations above spontaneous activity method, which computes the stimulus level for which the median spontaneous firing rate is 95% certain to be different from the threshold stimulus-induced firing rate (Raggio and Schreiner, 1994; Moxon, 1974, Witte et al., 1998; Sumner et al., 2005).

Using our SS-GLM statistical framework, we propose to compare the confidence bounds of the stimulus component in each time $b\Delta$ of the baseline response to the stimulus-induced response. While many methods of determining threshold are possible, we follow a concept recently introduced in behavioral data analyses (Smith et al., 2004) and define the MTML detection threshold as the lowest current stimulation level, v, at which there is reasonable certainty (> 0.95) that the electrically-induced spiking responses for that stimulus level are statistically different than spontaneous activity, and remain so for higher current levels. We write different, rather than greater, to allow for the possibility that the instantaneous rate of firing in a given bin will actually be smaller than the spontaneous activity with increased stimulus level, if the stimulus had a suppressive effect on the neuron.

To formally introduce our threshold inferences, we introduce a random variable s which represents a stimulus and takes on the values from $1, \dots, V$, and further denote the specific stimulus level as $s_{v'}$. In finding an accurate encoding for the spike trains, we estimate the likelihood $L(s_{v'}|N_{1:V}) = \Pr(N_{1:V}|s_{v'})$

$$
\begin{aligned}
L(s_{v'}|N_{1:V}) = \prod_{v=1}^{V}\prod_{k=1}^{K}\prod_{b=1}^{B}\left[\lambda\left(b\Delta|H_{v,k,b}, \widehat{\Psi}_{v'}\right)\Delta\right]^{n_{v,k,b}} \\
* \left[1 - \lambda\left(b\Delta|H_{v,k,b}, \widehat{\Psi}_{v'}\right)\Delta\right]^{1-n_{v,k,b}}
\end{aligned}
\tag{2.12}
$$

as a statistical characterization of the spike trains to stimulus $s_{v'}$, where $\widehat{\Psi}_{v'}$ is the ML estimate

27

of the encoding parameters for responses to stimulus level $s_{v'}$. To find the probability that a spike train $n_{v,k}$ came from actual stimulus level $s_{v'}$, we use Bayes theorem to write:

$$\Pr(s_{v'}|n_{v,k}) \propto \Pr(s_{v'}) \Pr(n_{v,k}|s_{v'}),$$

(2.13)

where

$$P(n_{v,k}|s_{v'}) = \prod_{b=1}^{B} [\lambda(b\Delta|H_{v,k,b}, \widehat{\Psi}_{v'})\Delta]^{n_{v,k,b}}$$
$$* [1 - \lambda(b\Delta|H_{v,k,b}, \widehat{\Psi}_{v'})\Delta]^{1-n_{v,k,b}}.$$

(2.14)

If we assume that the stimulus levels are presented equally likely (i.e. the distribution of $\Pr(s_{v'})$ is uniform), then:

$$\Pr(s_{v'}|n_{v,k}) \propto \Pr(n_{v,k}|s_{v'}).$$

(2.15)

Therefore, we find the first stimulus level for which, with probability $\Pr(s_{v'} > s_1|N_{1:v})_{b=b'}$ greater than 95%, the response at time $b'\Delta$ to stimulus level $s_{v'}$ is different from the response at the baseline stimulus level s_1. We make the inference based on the portion of the response attributable to the effect of the stimulus, while controlling for the effect of the biophysical properties of the neuron. We select the MTML threshold as the stimulus level for which the confidence bounds of the stimulus effect in at least one bin of the stimulus effect do not overlap and remain separate as the stimulus level continues to increase.

3.5. *Threshold estimation based on discrimination between the spontaneous activity and another stimulus level using a single trial*

Rather than inferring threshold using multiple trials and levels, one may want estimate neural discrimination on a trial-by-trial basis. For example, this type of decoding has been performed using SDT to explore whether the response of one neuron alone is sufficient to predict behavioral performance (Britten et al., 1992). In this decoding analysis, we estimate the probability that given two spike trains, one from a baseline stimulus level and one in response to a stimulus, the encoding model would correctly discriminate between the two stimulus levels.

28

This decoding analysis is easily performed in our framework using the ML estimates of the encoding parameters, $\hat{\Psi} = (\hat{\theta}_0, \hat{\gamma}, \hat{\Sigma})$. Given a spike train $n_{v,k}$, where v is any stimulus level from baseline to V and k is the trial, we evaluate the likelihood that the spike train came from some stimulus level s_v in our encoding model using Eq. (2.14). We can calculate the vector of likelihoods $L^T(s_{1:V}|n_{v,k}) = [L(s_1|n_{v,k}), L(s_2|n_{v,k}), \dots, L(s_V|n_{v,k})]$ as a function of stimulus level for all levels $v = 1, .., V$ in the encoding model. Then the maximum of the likelihood $L_{max}^T(s_{1:V}|n_{v,k}) = \max\left(L^T(s_{1:V}|n_{v,k})\right)$ is reached at the stimulus levels (from 1 to V) that makes the spike train $n_{v,k}$ most likely. Further, the concavity of the likelihood vector $L^T(s_{1:V}|n_{v,k})$ around the most likely stimulus level indicates the certainty with which the response can be attributed to that stimulus level. The value of $L_{max}^T(s_{2:V}|n_{v,k})$, i.e. the likelihood of the most likely stimulus level compared to the baseline-level likelihood, $L(s_1|n_{v,k})$ indicates the certainty that the spike train originated from a stimulus level above baseline using the information from a single spike train $n_{v,k}$.

We use the above defined likelihoods to suggest an SS-GLM based metric of discrimination between a response $n_{v,k}$ at stimulus level v from baseline activity $n_{1,k}$ to measure *the proportion of trains correctly discriminated from baseline with 95% certainty*. This metric can be calculated as follows:

1) Calculate the distribution of likelihood ratios for the response to baseline level ($v = 1$) for all trials $k = 1, \dots, K$ in the dataset in the following steps:

    a) Find the likelihood that the baseline activity arose from no stimulation, $L(s_1|n_{1,k})$.

    b) Calculate the vector of likelihoods, $L^T(s_{2:V}|n_{1,k})$, that the baseline activity resulted from any stimulation levels $s_v = 2, \dots, V$ above baseline, and find the likelihood at the most likely stimulus, $L_{max}^T(s_{2:V}|n_{1,k})$

    c) Compute the baseline response likelihood ratio $LR_b = \dfrac{L_{max}^T(s_{2:V}|n_{1,k})}{L_{max}^T(s_{2:V}|n_{1,k}) + L(s_1|n_{1,k})}$ for trial k

d) Repeat a) through c) for trials $k = 1, ..., K$

e) Construct the 95% confidence interval for LR_b such that the lower (upper) bound is equal to 2.5 (97.5) percentile of K values from c).

2) For each stimulus level v above baseline (i.e. for each $(2 \leq v \leq V)$), calculate the distribution of likelihood ratios for the responses originating from levels above baseline in the following steps:

a) For trials $k = 1, ..., K$ in the dataset, find the likelihood that the stimulus-induced response originated from the baseline level, $L(s_1|n_{v,k})$.

b) Calculate the vector of likelihoods, $L^T(s_{2:V}|n_{v,k})$, that the stimulus-induced activity resulted from all stimulation levels $s_v = 2, ..., V$ above baseline, and find the likelihood at the most likely stimulation level, $L^T_{max}(s_{2:V}|n_{v,k})$.

c) Compute the stimulus-induced likelihood ratio $LR_{SI} = \frac{L^T_{max}(s_{2:V}|n_{v,k})}{L^T_{max}(s_{2:V}|n_{v,k}) + L(s_1|n_{v,k})}$ for trial k.

d) If the stimulus-induced likelihood ratio, LR_{SI}, falls outside the 95% confidence interval of the likelihood ratio for baseline stimulation, LR_b, this response is classified as being correctly detected as different from baseline.

e) For the given stimulus level v, we compute the number of correct classifications all trials $k = 1, ..., K$.

Our proposed SS-GLM classification can be further used in the SDT framework described in the next text. Current methods of discrimination typically use SDT to measure the trial-by-trial discrimination performance. The SDT-based receiver operator characteristic analysis is a common tool to select optimal models from suboptimal ones without specifying the model class (Green and Swets, 1966) within the hypothesis testing framework. The receiver operator characteristic, ROC, trades the correct classifications (or true positives, i.e. stimulus at a trial is correctly classified as the above baseline stimulation) versus incorrect classifications (or false positive, i.e. stimulus at a trial is incorrectly classified as above baseline stimulation) as

30

the criterion used to differentiate between the two groups changes. The criterion may be firing rate (Britten et al., 1992; Shackleton et al., 2003), phase locking measuring the degree of temporal precision of firing (Lane and Delgutte, 2005), or another metric. We propose a criterion based on our SS-GLM methodology and compare it to a commonly used rate-based SDT analysis (see RESULTS section 4.5) which was also applied by Lim and Anderson (2007). Furthermore, Lim and Anderson use the same dataset used here. To do so, we repeat their analysis of computing the ROC curve, which estimates the proportion of trials for which the number of spikes in each trial exceeds a criterion spike count against the proportion of trials for which the criterion is exceeded during each baseline trial for different spike counts. In addition, we recomputed the ROC area under curve (AUC) which is equal to the probability that the classifier will assign a higher score to the true positive (stimulus-induced level) rather than the false positive (baseline level) and is equivalent to the Mann-Whitney U test as well as the Wilcoxon test of ranks (Fawcett, 2006). Following Lim and Anderson's (2007) rate-based analysis, we compute the ROC AUC based on the number of spikes in each trial during the first 28 msec following the pulse. We compare this discrimination performance using our metric of the stimulus level that makes the likelihood of the spike train arising from a stimulus-induced level, $\max L^T(s_{2:V}|n_{v,k})$, most likely to the most likely stimulus level to elicit the baseline data. We compute the discrimination performance in the SDT framework using out metric, to directly compare our results to rate-based metrics.

3.6.   *Measuring discrimination performance between a given stimulus level and a baseline condition, using information from multiple trials*

The analyses mentioned in section 3.5 can be extended to measuring discrimination performance over multiple trials from a given stimulus level v. To extend the methodology, we can calculate the likelihood of observing all the trials from level v, assuming the parameter estimates from stimulus level v':

$$L(s_{v'}|n_{v,1:K^*}) = \prod_{k=1}^{K^*}\prod_{b=1}^{B}[\lambda(b\Delta|H_{v,k,b}, \widehat{\Psi}_{v'})\Delta]^{n_{v,k,b}}[1 - \lambda(b\Delta|H_{v,k,b}, \widehat{\Psi}_{v'})\Delta]^{1-n_{v,k,b}} \tag{2.16}$$

The likelihood computed over multiple trials in Eq. (4.5) compared to the most likely stimulus level to have elicited the baseline response over the multiple trials of interest. These most likely levels can then be used as metrics in the ROC AUC calculations and the likelihoods can be used to calculate the likelihood ratios used to measure percent correct discrimination with 95% confidence, similarly to the procedure described in section 3.5.

3.7. Decoding across multiple neurons

The decoding analysis comparing performance for a single trial at two stimulus levels can be expanded to combine spiking activity from several neurons and thus to define a more realistic statistical population model of neural decoding. It is quite unrealistic to assume that only one neuron participates in the decoding of spike trains for each stimulus level. More likely is the idea that the responses of multiple neurons are combined to decode the single trial of stimulus and compare the responses between two different levels. Our framework can directly be extended for several neurons. It assumes that ML estimates exist for models of responses for neurons $c = 1, ..., C$. Applying the ML principles from the statistical methods for analysis of neuronal interconnections (Borisyuk et al., 1985; Brillinger 1988, 1992; Chornoboy et al., 1988; Okatan et al., 2005), we directly extend the definition of our conditional intensity function by including the dependence on firing behavior of other neurons. Thus the likelihood of a specific stimulus $s_{v'}$ is:

$$L(s_{v'}|n_{1:C,v,k}) = \prod_{c=1}^{C}\prod_{b=1}^{B}[\lambda(b\Delta|H_{c,v,k,b}, \widehat{\Psi}_{c,v'})\Delta]^{n_{c,v,k,b}}$$
$$* [1 - \lambda(b\Delta|H_{c,v,k,b}, \widehat{\Psi}_{c,v'})\Delta]^{1-n_{c,v,k,b}} \tag{2.17}$$

We can then do the same decoding calculations as those based on the vector of likelihoods for a single neuron in sections 3.4-3.6. This formulation of the likelihood in Eq. (2.17) does not

assume that the neurons are independent, since the definition of the conditional intensity function for each neuron may involve interactions with nearby neurons via e.g. the dependence on spiking history of other neurons (Truccolo et al., 2005). Our framework thus can be directly extended to combine information across several neurons to make predictive statements about overall system threshold-detection performance.

3.8 Neural data description

To illustrate our methodology of neural inferences, we analyze 30 multi-unit recordings (MU) from the primary auditory cortex in response to electrical stimulation of the IC at 19 stimulus levels across eight anesthetized guinea pigs (Lim and Anderson, 2007). Multi-site silicon substrate probes were used, each with 16 sites (NeuroNexus Technologies, Ann Arbor, MI), for stimulation and recording,. For each probe pair, only one site within a certain best frequency region within the inferior colliculus was stimulated while the corresponding neural activity recorded on the site within the main input layer of the primary auditory cortex with a similar best frequency was used for analysis. Methods for probe placement and further description of the neural data are presented (Lim and Anderson, 2006; 2007).

The stimulus consisted of a single-biphasic, charge balanced current pulse (200 μs/phase, cathodic leading) from 1 to 56.2 μA presented in logarithmic steps at 2 dB increments relative to 1 μA. Forty trials were recorded for each stimulus, including 40 trials of spontaneous activity. The stimuli were presented at a rate of two per second and randomized across all stimulus levels. The response was recorded from 10 ms before to 50 ms after the stimulus onset. Spike detection was performed by taking any value of the recorded voltage signal 3.5 times the standard deviation of the background (without spikes) signal and time-stamping the largest positive or negative deflection as a spike. To assure that one spike was not being detected twice, a 1 ms lock-out period was employed. This lockout window also ensured that no two spikes occurred within 1 ms of one another. The lockout window was necessary because of the difficulty in isolating single units in this preparation (Lim and Anderson, 2007).

33

4 RESULTS

4.1 Building an SS-GLM model to characterize neural responses

In order to illustrate our encoding methodology, we analyze a response from one of our set of 30 MUs in the primary auditory cortex of the guinea pig. As the dot raster of this example response illustrates (Figure 1A), the response is recorded between 10 ms prior to and 50 ms after stimulus onset. One common question neurophysiologists ask is to characterize the changes in the raster plots across different stimulus levels. Looking at the response dynamics across stimulus levels, we visually observe that the response is similar to baseline until approximately stimulus level nine. At approximately the stimulus level 10, we begin to see some changes in the response several milliseconds after stimulus onset, which is clearly visible in the dot raster by stimulus level 12. At higher stimulus levels, the increased firing rate response occurs at longer time intervals. Using these subjective qualitative visual observations, we would estimate the visual threshold, the first level for which the responses are significantly different from baseline, anywhere from stimulus level 10 to level 12, depending on the criteria and subjective judgment of the experimenter.
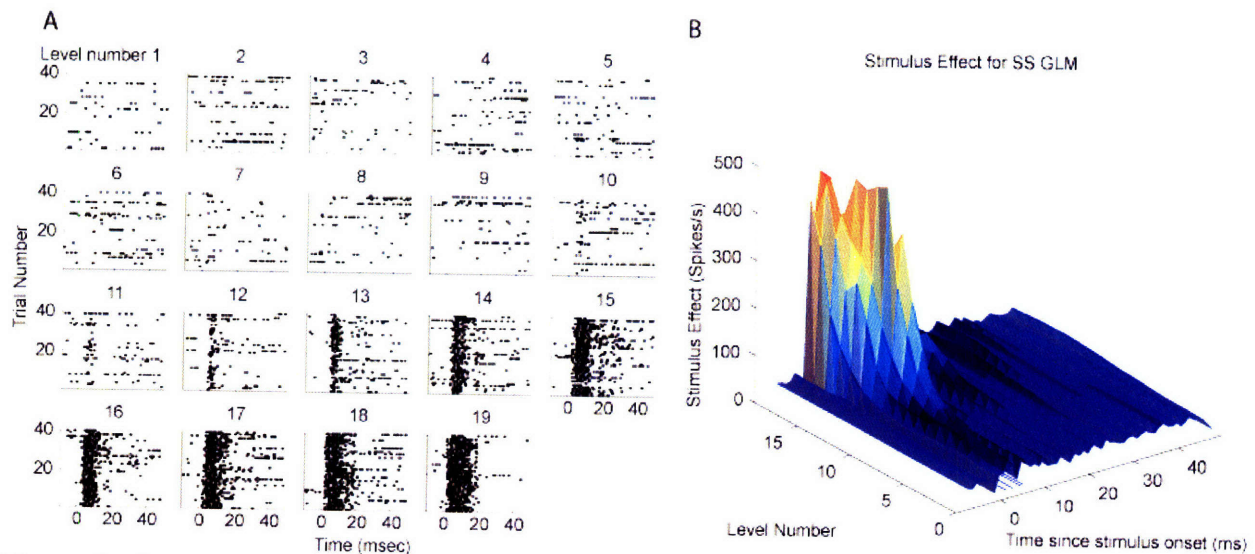
Figure 1: A. Raster plots for an example response of a neuron from the primary auditory cortex (A1) at 19 different stimulus levels imposed to the inferior colliculus (IC). The response to level 1 is the baseline response in the absence of electrical stimulation. The time 0 ms represents the time of the stimulus onset. The responses are recorded from 10 ms before to 50 ms after stimulus onset. The stimulus is presented 40 times per level and at 18 different stimulus levels from 1 μA

to 56.2 µA in 2 dB steps with an additional 40 trials of spontaneous activity (level number 1). B: Estimated stimulus effect (2.5) for the state-space generalized linear model based on the data in A.

A non-subjective way of characterizing the changes in neural responses across the stimulus levels is by calculating the statistical significance of the observed differences in neural responses. Before we make any inference from the data, however, we first build a model that appropriately describes the data. Our approach is to find a model that captures the entire response dynamics over the recording duration of all trials and levels. This approach is different from the commonly use rate-based models that, for this brief stimulus, estimate thresholds using only the neural response shortly after the stimulus onset. The duration of the onset response can vary with stimulation level and neuronal type, thereby making the choice of duration of the modeled response somewhat subjective. For example, when analyzing these data, Lim and Anderson (2007) used the first 28 ms following the stimulus onset to calculate rate-based threshold estimates.

To characterize the responses (Figure 1A), we fit the SS-GLM model (Eqs. (2.5) and (2.7)), where we assume that the firing response depends on the stimulus and spike history. The estimate of the stimulus effect (Figure 1B) illustrates how the stimulus effect changes as a function of stimulation level and of time since the stimulus onset. This stimulus effect captures such dynamics as an increase in firing between 0 and 20 ms with level and the increase in the duration of the increased firing rate after the stimulus onset with level. These characteristics are also seen in the raw data shown in the dot raster (Figure 1A). The second component of our SS-GLM model is the history of the current spiking response which describes the effect of the intrinsic biophysical properties of the neuron. Because the recording duration is itself so short (60 ms per trial), we select a short duration of history since our choice of history duration directly affects the total size of the response that we can model. With a choice of 4 ms, we can model the remaining 56ms of the response dynamics. We cannot model the first 4 ms of the response, because the past spiking history for that portion of the response is unknown. We

35

point out here that for this dataset, we implemented a 1 ms lockout window ensuring that the point process assumption of binary data holds.

4.2   The SS-GLM model offers a greatly improved model fit over rate-based methods

We verify the underlying fit of our model to the data by using goodness-of-fit analyses suitable for assessing point process models. First, we assess that the rescaled intervals (Eq. (2.10) derived from our estimated intensity function are uniformly distributed by looking at the Kolmogorov-Smirnov plot. The KS plot (Figure 2A, black line) fits closely to the 45-degree line and virtually within the 95% confidence bounds of the KS statistic indicating that the distribution of the rescaled intervals is close to uniform within the bounds of the confidence intervals. Second, to assess that the spike times are independent, we rescale the estimated spike times from the instantaneous firing intensity function to a Gaussian distribution and examine their autocorrelation. The autocorrelation of the Gaussian rescaled intervals (Figure 2B) are mostly within the 95% confidence bounds and only show some correlation for early lags. This residual correlation is most likely due to the short duration of history chosen in this model. Had the data included baseline activity for longer durations prior to stimulus onset, the modeled history duration could have been increased and the residual autocorrelation likely reduced (Czanner et al., 2008). Both of these statistics indicate that the SS-GLM model is an adequate description of the spiking dynamics, although the fit could be improved with a longer recording duration allowing for models of longer history dependence.
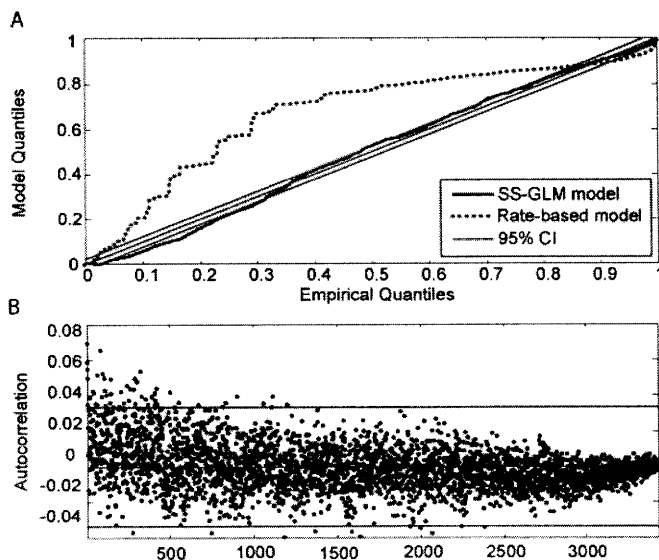
Figure 2: Goodness-of-fit assessment for state-space generalized linear model and rate-based model for the data in Fig 1A. A: Kolmogorov-Smirnov plot for SS-GLM model (black solid) and for rate-based model (gray dotted) for which a single firing rate is computed per level. Gray solid lines indicate 95% confidence intervals of the KS statistic. B: Autocorrelation of the Gaussian rescaled intervals and 95% confidence intervals.

In contrast to the adequate model fit of the SS-GLM, the underlying model fit to the rate-based model is poor (Figure 2A). The KS analysis is performed by assuming a constant rate over the duration of the trial, estimated as the total number of spikes per trial. Only the response during the first 28 ms since stimulus onset is used to compute the goodness of fit for this rate-based model, following previous analysis (Lim and Anderson, 2006) to reflect the data used in the threshold estimate calculations. The KS plot implies that this rate-based model does not adequately capture the stochastic structure of the data. Dot raster plots in Figure 1 show that, particularly for lower levels, the firing rate is not constant, even for the first 28 ms of the response since stimulus onset. Because the goodness-of-fit of our models suggests that the models capture the stochastic nature of the data well while those of the rate-based models are less suggestive, we have increased confidence in our threshold estimates using our SS-GLM model.

4.3   Threshold estimate based on multiple trials at multiple stimulus levels

In our framework, we estimate the effect of the stimulus while controlling for the effect of the history that allows us to make inferences about how the portion of the response attributable to the stimulus changes with increased levels of stimulation. Our threshold inference based on all the data (Figure 3) is made by comparing the stimulus effect at baseline to the stimulus effect at higher levels by using confidence bounds (95th percentile) of the stimulus effect. We propose to estimate the threshold as the stimulus level for which the confidence interval of the response in at least one temporal bin of the stimulus effect does not overlap with the confidence interval for that bin at the baseline level and continues not to overlap for at least that bin for all higher levels. Our definition for this MTML threshold is purposefully broad. It allows for instances when the stimulus has a purely suppressive effect on the response, such that the response falls below baseline activity for higher levels. For the example in Figure 3, stimulus level 8 is chosen as threshold because the lower confidence bound of the instantaneous firing rate in the 6 ms bin from stimulus onset exceeds the higher confidence bound of the

spontaneous rate and the confidence bounds do not overlap for all consequent levels for that time bin.
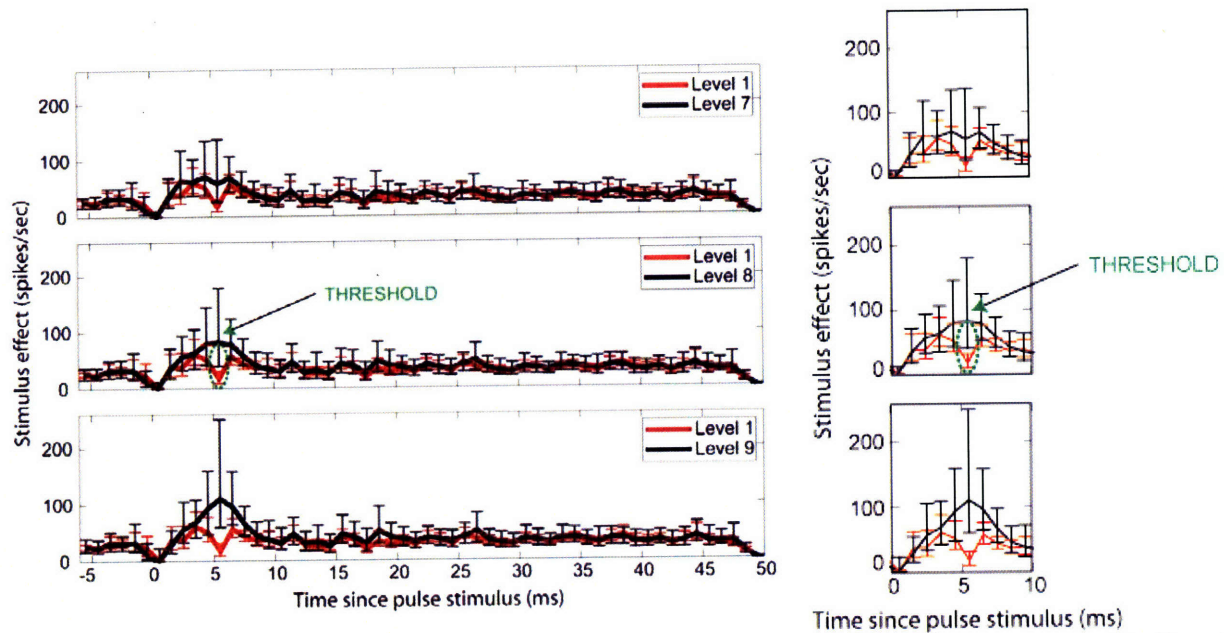
Figure 3: Illustration of MTML neural threshold response characterization using SS-GLM. For all levels below threshold, the confidence bounds (95th percentile) of the stimulus component for the baseline level and sub-threshold levels overlap for all time bins (top). At threshold (middle), the confidence bound of the stimulus component for at least one time bin does not overlap with the confidence bound at baseline and remains as such for that time bin for higher levels (bottom). Left figure shows entire modeled duration and right figure enlarges the times during which threshold is isolated.

Using the visual method, the dot raster of responses across all trials and levels is examined to find the first stimulus level for which the response begins to differ from baseline at any point in the recording window. Our first threshold definition offers a mathematically based alternative to the subjective visual method. Our threshold definition can also be compared to the RLF method, although less directly. In the RLF method, the total spike rate is taken across a set time window and threshold is estimated as the stimulus level for which the normalized spike rate reaches some percentage of the maximum driven rate (for example, 25% -- Bierer and Middlebrooks, 2002, although some use 10%). Some define the MTML threshold as the stimulus level for which the median spontaneous and stimulus induced firing rates differ with 95% confidence (Witte et al., 1998; Sumner et al., 2005). We refer to this later method as the standard deviations above spontaneous activity (SDASA) method. Although the later two

methods are rate based, the similarity to our methodology is that the data across all trials and levels are used to make the threshold inference. Of the two model-based methods, the SDASA method is more similar to our method since both detect threshold using a 95% level of certainty that the baseline activity differs from the stimulus induced activity at that level. For the non-subjective methods, given our assumptions about the rate-window, the threshold estimates are stimulus level 11 and level 12 for the RLF method (using the Bier and Middlebrooks, 2002 criteria) and SDASA method, respectively, compared to our estimate of level 8. Each level step represents a 2 dB increase in current level, indicating that for this example, the SS-GLM estimates threshold 6-8 dB lower than the rate-based methods partly due to the additional temporal information extracted from the spiking activity. Across the entire dataset of 30 MUs, our threshold estimates differ from the RLF and SDASA methods by 8 dB and 6 dB, respectively.

4.4 Threshold estimates based on discrimination between baseline and another stimulus level using data from one or multiple trials

We have described our methodology for computing the threshold inference using all the data in the experiment. One may instead want to assess the ability of the neuron to discriminate between two levels on a trial-by-trial basis. Since this inference is based on less information than the MTML threshold inference, one would thus expect decreased sensitivity, and therefore higher thresholds than those based on all the data. Another assessment of discrimination is regarding the ability of the neuron to discriminate between two levels based on multiple trials. This is the intermediate case between discrimination based on responses from multiple trials at multiple levels and a single trial from two different levels, and would thus be expected to yield higher thresholds than the first inference and lower thresholds than the second inference (see Eq. (2.16)).

Likelihood-based decoding for our framework is illustrated in Figure 4. The top row shows the vector of likelihoods as a function of level for a validation set of 15 neural responses

from each stimulus level. Each likelihood represents the probability that the trial from the actual level (x-axis) arose from stimulation at the potential level (y-axis). Light areas represent high probability and dark areas represent low probabilities. Likelihoods are computed for data consisting of a single trial from one level to 40 trials at one level. The encoding models used to calculate these likelihoods are based on a fitting dataset of a separate 25 trials from each stimulation level. Using a single trial, the likelihood that the response arose from one of the first 11 or 12 levels of stimulation is approximately equally high (Figure 4 top left). For levels above 11 or 12, we can assign the response to approximately the correct level, although the precision is poor (Figure 4 top left). However, when the decoding scheme is able to combine information across multiple trials across each level (rightmost top figure), discrimination improves. The decoding scheme is now able to discriminate between baseline and higher levels earlier (approximately around level 10 or 11 for 40 trials) and with greater precision. If the decoding scheme is also able to combine information across all levels, the discrimination approach that using all levels and all trials simultaneously (Figure 3).
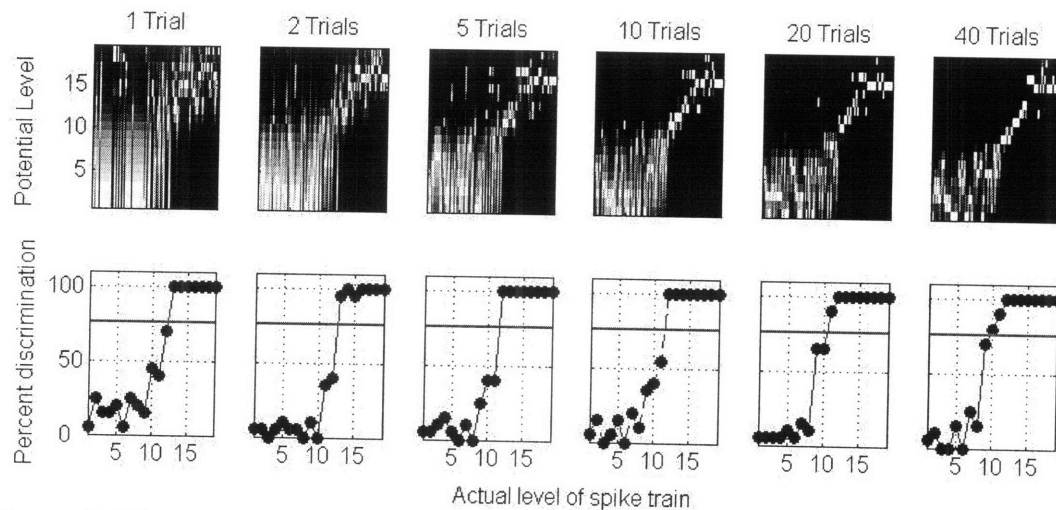
Figure 4: Illustration of threshold estimation using SS-GLM decoding using information from one (leftmost column) or multiple (all but leftmost column) trials comparing responses at two levels (stimulus-induced and baseline). To draw the analogy to signal detection theory based methods, we defined the threshold as the level for which the proportion of correct responses exceeded 76% (shown by gray line). A. Likelihood, as a function of level, that a spike train(s) arising from some actual stimulation level (x-axis) came from the each potential level (y-axis), using the maximum likelihood coefficients estimated in the modeling encoding. Encoding model estimated using 25 trials from each level and likelihoods computed in decoding analysis are based on 15 other levels

in the validation dataset. The number of trials at the top of each plot indicates the number of random trials chosen from the validation dataset with replacement. Light areas represent high probability and dark areas represent low probability. B. Metric for describing these likelihood matrices as percent of trials (or groups of trials) for which the likelihood ratio from a stimulus-induced trial exceeded the 95th percentile of the likelihood ratio from the baseline trial (or groups of trials). (For more details on this analysis, see Methods, section 3.5)

Although the likelihood-based decoding shown in Figure 4 (top) may be sufficient, we can summarize these results by plotting the proportion of spike trains that are correctly discriminated from baseline with 95% certainty. This metric uses the concavity of the likelihood vector as a measure of certainty to calculate this metric (see Methods section 3.5) and we plot the results in Figure 4 bottom rows. To compare the metric across combinations of multiple trials, we draw a line at 76%, indicating that the level for which the metric intersects this line corresponds to the level at which we can be 95% certain that the trials from this level can be correctly discriminated from baseline. Looking across the metric as the number of trials increases, we notice that the level for which 76% correct performance is 95% certain shifts toward lower levels, increasing from level 13 for using 1 trial to level 10 combining across 40 trials. Nevertheless, when combining across as many as 40 multiple trials, the discrimination between baseline and levels up to level 8 is uniformly poor, indicating that the information to discriminate between these levels is not sufficiently present. These observations mirror the observations that can be made from the confusion matrices in Figure 4 top rows.

4.5 Comparing SS-GLM threshold estimates to those using rate-based Signal Detection Theory

The SDT-based analysis is often used to predict performance for one neuron on a single trial basis. We compare the prediction power of our methods on a single trial basis to those of rate-based SDT for all 30 MUs in our dataset. The assumption underlying the rate-based SDT method is that the number of spikes around the onset of the stimulus is a sufficient statistic to predict single-trial performance. To compare between the two methods, we compute the encoding model in our framework and decode the most likely stimulus level to elicit a spike

train. We then use that most likely level in the SDT framework to predict single-trial performance.

We compare our methods to those using SDT and illustrate the results for two example MUs are presented in Figure 5A1 and B1. The response in Figure 5A1 is the example response shown in Figure 1, while the response in Figure 5B1 is an additional response from the dataset. To compare our discrimination method to the rate-based SDT method, we perform 35 experiments for each of two example MUs. For each experiment, we select a random 25 trials for each stimulus level for the training dataset and leave the remaining 15 trials for the validation dataset. We fit the SS-GLM model on 25 trials of the training data. For our discrimination method, we perform a likelihood decoding analysis to select the ML stimulus level corresponding to each spike train in the 15 trials of the validation dataset. We compute the ROC comparing the ML levels for each induced stimulus level and the ML stimulus level of the spontaneous activity (Figure 5 A2 and B2) and compute the area under the ROC curve for each stimulus level, defined as the area under the curve (AUC) statistic, by summing under the ROC curve. To compare our method to the rate-based SDT method, we also compute the ROC AUC statistic based on the number of spikes occurring for each trial and stimulus level in the 15 trials of the validation dataset between stimulus onset and 28 ms after onset (as described in Lim and Anderson, 2006). Although when calculating the discrimination performance for our method, we used the entire recording duration, we limited the duration when calculating the AUC statistic based on spike number to be consistent with current approaches of only using the temporally-limited response. Current approaches limit the duration of analyzed firing rates using the rate-based SDT method because the number of spikes is averaged only over the temporally-limited response, where the firing rate tends to be higher. Including the same duration of data using the rate-based SDT approach as we had included in our modeling effort would degrade their estimates to increase the estimated threshold level. Lastly, we calculate the difference in the ROC AUC obtained using our method and the current rate-based method (Figure 5 A3 and B3).

42

The AUC statistic is asymptotically equivalent to the proportion of trials originating from a stimulus level above baseline that can be correctly discriminated (given a higher score) from trials originating from baseline. We repeat the analysis for all 30 MUs in the dataset, and find that for 70% of the units (21 out of 30 MUs), the median of the difference between the two AUC metrics is greater than zero in the region of single-trial threshold based on the rate-based SDT method. This finding indicates that the use of the SS-GLM encoding model to decode threshold improves threshold estimates based on a single-trial (p<.05).

The SS-GLM model outperforms the rate-based SDT method for stimulation levels where the response exhibited a strong onset with few spikes (Figure 5 A1 levels 8-12, B1 levels 9-12). This improvement for the SS-GLM model is shown for two simulations for each MU (Figure 5A2 and B2) with the AUC curve based on SS-GLM lying above the AUC curve based on number of spikes. Using number of spikes model, the threshold is often taken as the stimulus level for which ROC AUC reaches 76%. In these examples, the model based on SS-GLM would yield a threshold estimate one stimulus level number lower than the estimate based on number of spikes.
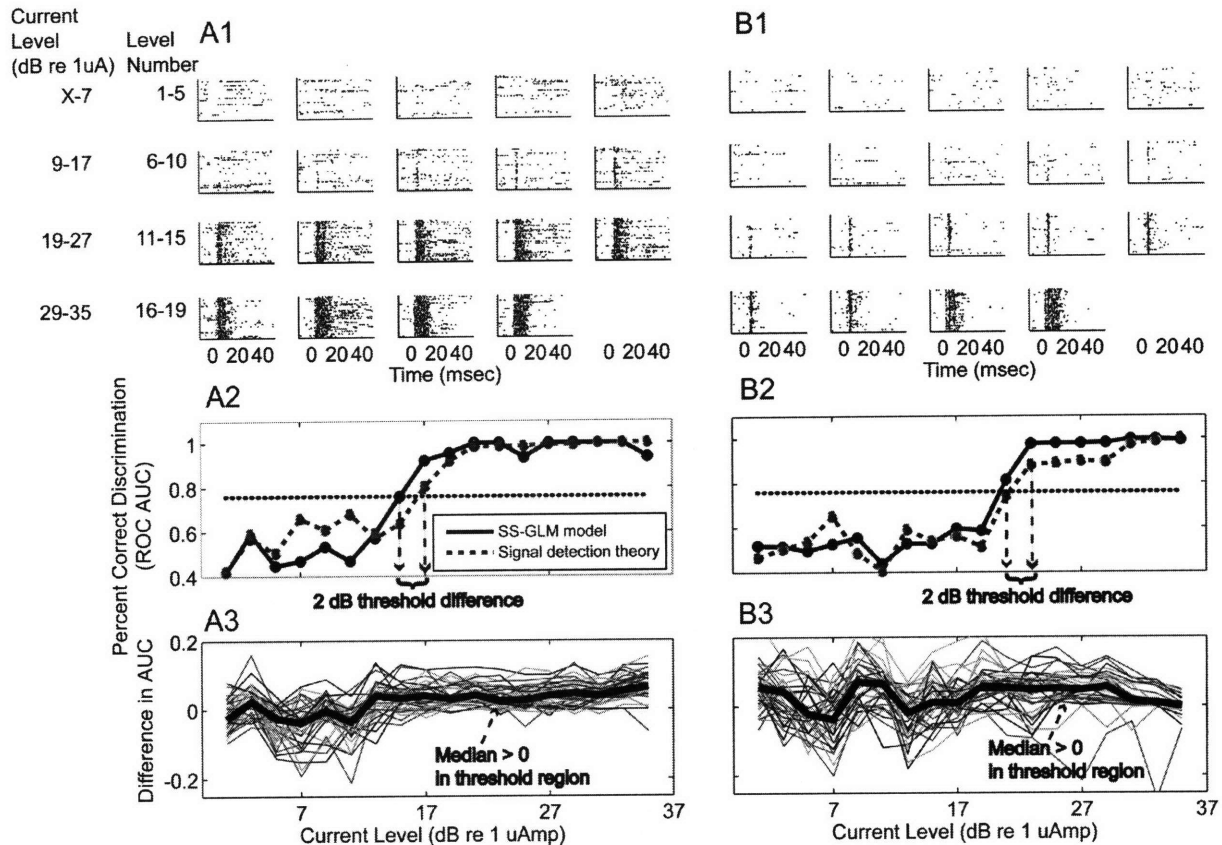
Figure 5: Comparison of the area under the receiver-operator characteristic curve measurement using the state-space generalized linear model with that using the traditional number of spikes per trial metric for two multi unit recordings in our dataset. For each multi-unit recording (A1 and B1), a random 25 trials are selected for each level and the state-space generalized linear model is fit 35 times, each time selecting different random trials in the fit datasets. The remaining 15 trials in each of the 35 draws are used in the validation dataset. For each curve shown in A2 and B2, the area under the curve is computed by comparing the maximum likelihood levels assigned to spontaneous activity and to stimulus induced trials (dotted lines A2 and B2) for all combinations of baseline and stimulus-induced trials for each of the 35 models. These curves are compared with traditional area under curve metrics based on number of spikes (solid curves A2 and B2). The spike number method used a recording window from 0 to 28 ms after stimulus onset, consistent with Lim and Anderson (2007). The SS-GLM model used the entire recorded duration. A3 and B3: Differences in area under curve metric between the state-space generalized linear model and rate-based method for 35 different fitted and validation datasets for each multi-unit cluster.

There are two main reasons why the SS-GLM-based discrimination outperforms the rate-based SDT method. First, the SS-GLM is able to capture the timing of the synchronized onset response, even in the absence of vigorous firing after the initial onset, while the rate-based SDT method can mistake this onset response for spontaneous firing with the same number of spikes. Second, the SS-GLM method provides a performance advantage for neurons with a high baseline rate (Figure 5A1), where the number of baseline spikes for some

44

trials is equal to the number of spikes in response to the stimulus. In this second scenario, the high baseline rate would decrease the value of the rate-based ROC AUC statistic at higher levels relative to the SS-GLM method.

4.6 Discrimination performance across multiple neurons

Behavioral detection depends not on the firing patterns of a single neuron responding to a single stimulus presentation, nor on a single neuron responding to multiple trials of one stimulus presentation. Instead, it depends on many neurons responding to a single presentation of the stimulus. In the following section, we illustrate how our framework can be extended to suggest a statistical model for this scenario to predict behavioral performance.

First, we show individual neuron decoding results for the 30 neurons in our dataset. The methods used here are those used in Figure 5A3 and B3. We conduct the decoding analysis by fitting the model to 25 randomly selected trials and leaving the remaining 15 trials in the validation set. We repeat the fitting 35 times for different fitting and validation sets each time. Using the validation dataset, we compute the proportion of trials for which the likelihood of a trial originating from a stimulus-induced level exceeded the likelihood of the trial that originating from the baseline level. These results are shown in Figure 6.

There is a wide range of discrimination performance over all the units in the dataset, as shown in Figure 6. Using the common metric of adequate detection as 76% correct detection (indicated in the figures as gray dotted lines), we observe that several units can discriminate, on average, between the no stimulus condition and a stimulus present condition as early as level 8 with 76% accuracy, while several cannot perform the task, on average, with this level of accuracy until level 16. Many neurons are average performers, relative to the best and worst in the set, such that they can discriminate, on average, between no stimulus and a stimulus present with 76% accuracy by levels 11 or 12.
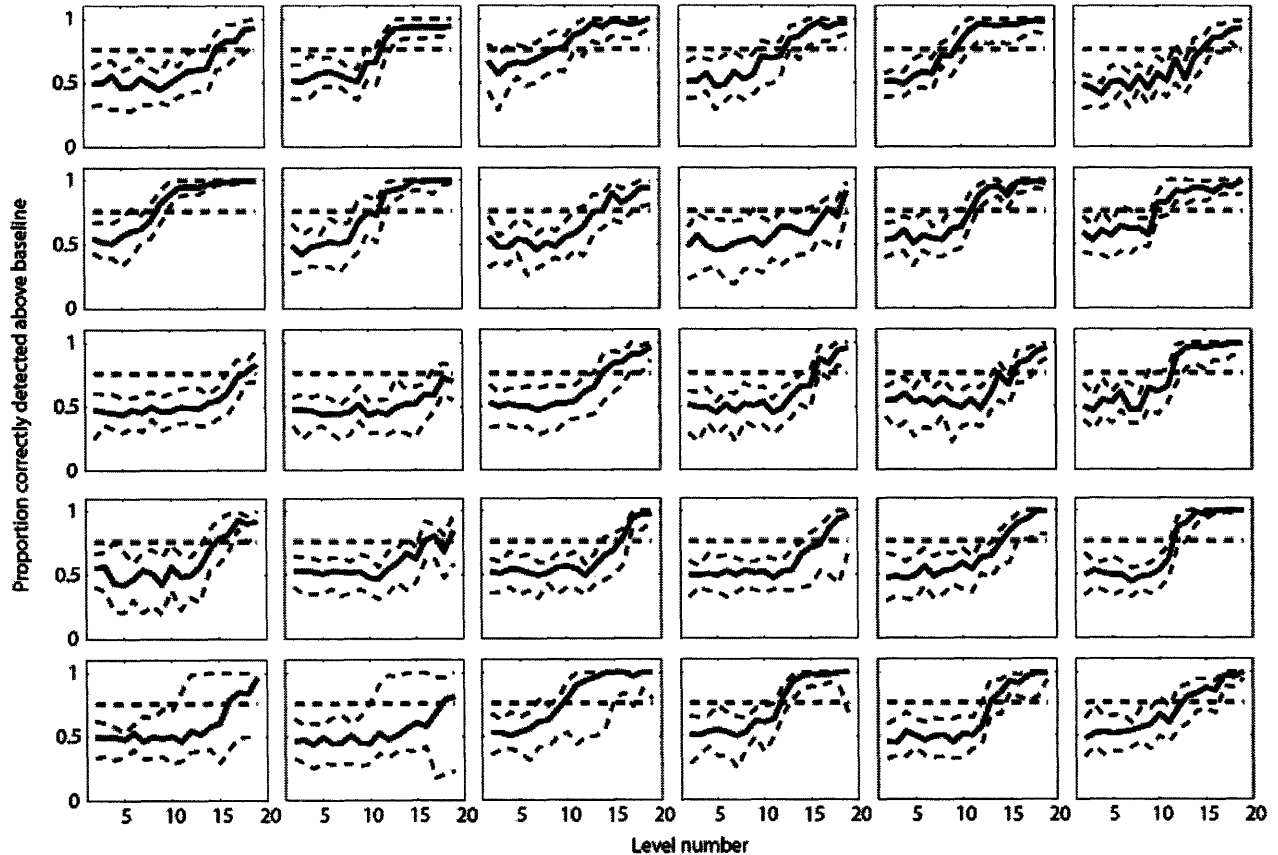
Figure 6: Decoding analysis performed for each of 30 neurons in dataset individually. Models are fit to 25 random trials and decoding is performed on remaining 15 trials in validation dataset. The process is repeated 35 times. See Figure 5 caption for methods. Thick black lines indicate median (50%) discrimination and dotted black lines indicate the 95% confidence bounds for these estimates. Gray dotted lines indicate 76% correct detection.

Figure 6 shows the range of discrimination performance for all the neurons in the dataset when the neurons discriminate individually, without input from other neurons. Next, we illustrate how responses across different neural types can be combined to model how, for example, an upstream neuron may discriminate based on inputs from several downstream neurons. Although our formulation does not preclude the introduction of neural interactions into the model, here, we illustrate a case where the independence of neurons are assumed (see Chapter 5 for examples of models with neural interaction). Our simulation is based on our model of six actual units from our dataset, three with lower thresholds and three with higher thresholds. We divide the responses over 40 trials in the dataset for the 6 units into 25 fitting trials and 15 validation trials. Again, using the validation dataset, we perform decoding analysis comparing the discrimination between baseline and a stimulus-induced response all six units.

46

Figure 7A1 and B1 shows that the units form two groups: those that have median single trial

decoding performance of 76% correct around level 12 and those with median single trial

decoding performance of 76% at levels 15 and higher, respectively. We refer to the former

group as units with a low single-trial decoding threshold (LOW) and those in the latter group as

having a high single-trial decoding threshold (HIGH).
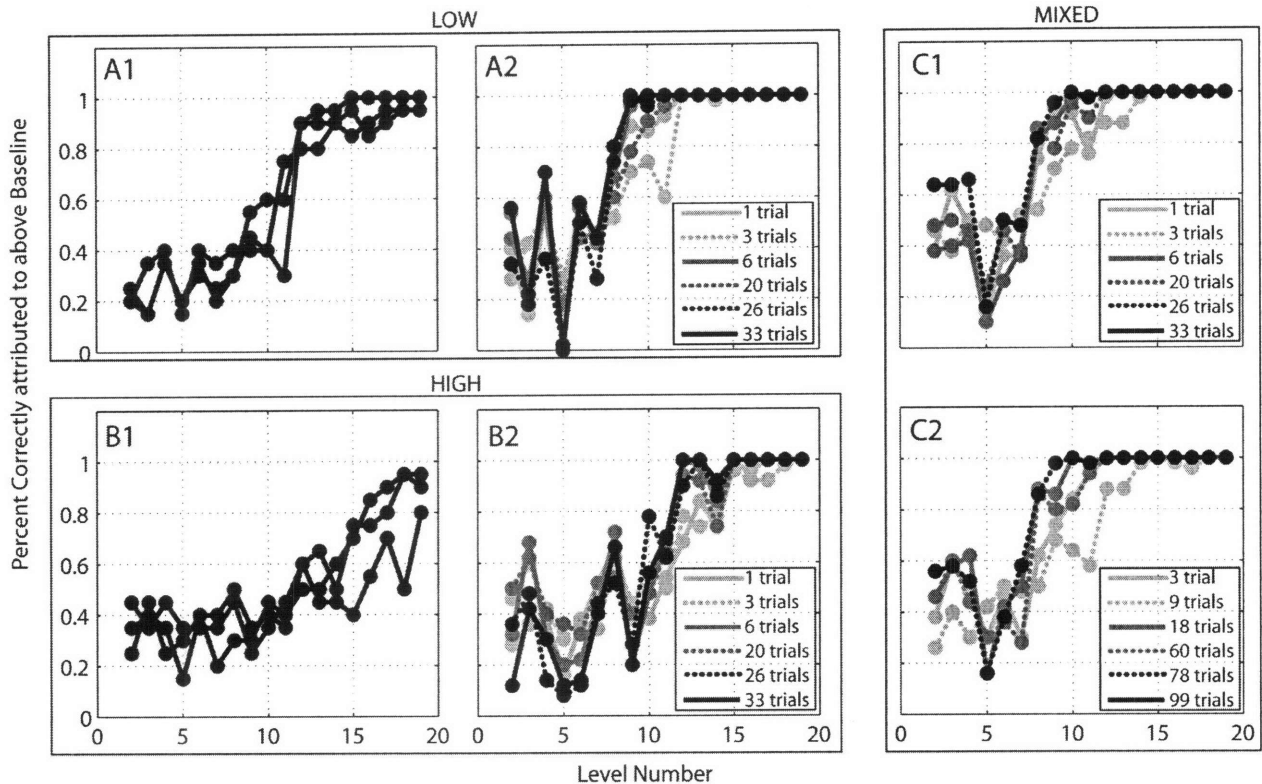
Figure 7: Illustration of decoding methodology for combining information across multiple neurons. All models are fit with 25 trials and decoding analysis is performed with the other 15 validation trials. Trials are taken from actual modeled responses having the single-trial decoding characteristics as shown in A1 and A2. Three neurons with single-trial decoding performance that reached near perfect detection around level 12 are used (A1, LOW) as well as three neurons that reached near perfect single-trial decoding performance around level 17 (B1, HIGH). Trials across the three low neurons (A2) and three high (B2) are combined for 1 (light gray dotted), 3 (light gray solid), 6 (gray dotted), 20 (gray solid), 26 (black dotted), 33 (black solid) trials from each neuron drawn randomly from the validation dataset and their likelihoods are combined. Trials from neurons in LOW and HIGH categories are combined by taking trials from 1 LOW and 1 HIGH neuron (C1). To compensate for 1/3 the number of LOW neurons used to compute the likelihoods in C1, the number of trials from each neuron is now multiplied by 3 (C2). See section 3.6 for detailed description of combining information across multiple neurons.

We treat the six units as six different classes of responses, such that each of the

validation trials can be thought of as a trial from a different neuron with the response

characteristics of one of the six units. In doing so, we simulate a scenario where the units in

each group are all contributing to the system decoding performance and examine how the system performance varies depending on the number of units from each group. First, we examine the scenario where the low and high units contribute separately to system performance. For each group, we examine how system discrimination changes when we vary the number of contributing neurons from 1 to 33 in each class of low or high threshold units, meaning that the number of contributing simulated neurons ranges from 3 to 99 for a single stimulus trial, since we have included 3 unit classes in each of the low and high subgroups. Figure 7A2 and B2 shows the results of the simulation for system discrimination, calculated as in (2.17) . For the low units, when combining across one neuron from each class (three neurons total), the discrimination performance is similar to that of each unit performing individually (compare to Figure 7A1 and B1). However, as the number of neurons increases, discrimination performance improves. For instance, for 1 neuron from each class (3 total neurons), the discrimination performance reaches 76% at level number 9 and close to perfect at level number 12, compared to reaching 76% at level number 7 and close to perfect performance at level number 9 for 33 neurons from each class (99 neurons total). A similar phenomenon occurs for high threshold neurons, with 1 neuron from each class (3 neurons total) showing similar performance to individual neuron performance (reaching 76% discrimination around level 14 and near perfect discrimination near level 18 or 19) while combining information across 33 neurons from each class (99 total neurons) improves discrimination (76% at level 12 and near perfect discrimination around level 14).

Next, we simulate a scenario where responses across low threshold and high threshold subgroups of neurons are combined. We conduct this simulation by combining responses from 1 low threshold neuron and 1 high threshold neurons for a varied number of single trials ranging from 1 to 33 from each of the 2 neuron classes (from 2 to 66 total trials). The discrimination performance is shown in Figure 7C1. Compared to discrimination performance of the low neurons only, the performance combining across high and low units is similarly good for high

48

levels of stimulation (level 8 and beyond), but poorer for low levels of stimulation (levels 2 through level 8). We hypothesized that the somewhat poorer performance is related to this mixed scenario having fewer trial contributions from low threshold units. Viewed from this perspective, if we multiply the number of trials by 3 to compensate for the fewer trials from each group of neurons, we see that the performance in the mixed case (Figure 7C2) is now similar to the performance for the low neurons only (Figure 7A2). These results agree with previous findings that the lower threshold unit contributions exceed those of high threshold units and that the number of lower threshold units is important in predicting system performance (Colburn, 1981; Delgutte, 1986; Winslow and Sachs, 1988).

As a final illustration of combining across multiple neurons, we calculate the discrimination performance across all 30 neurons in the dataset. Again, we assume that the neural responses are independent and are not a function of the interaction between neurons for simplicity. This result is shown in Figure 8. Again, 25 trials in the fitting dataset are used with the remaining 15 in the validation dataset and the model is fit 35 times for different choices of fitting and validation datasets. Decoding performance is computed as described in Section 3.6. For levels 2 through 7, the detection performance of the neuron set is approximately chance level. That is, the neuron set cannot discriminate between the no stimulus (level 1) and the stimulus present condition when the level is below level 7. However, the performance rapidly increases from level 8, reaching near perfect detection by level 11 and remaining so for all higher levels. The combined performance is on-par with the best neurons (for example, neuron 7 – second row first column Figure 6) but does not exceed the best performers. These results support those found when combining across different classes of neurons (Figure 7) that the overall combined performance is strongly influenced by the performance of the best neuron or neurons (Siebert, 1965; Parker and Newsome, 1998; also see discussion).
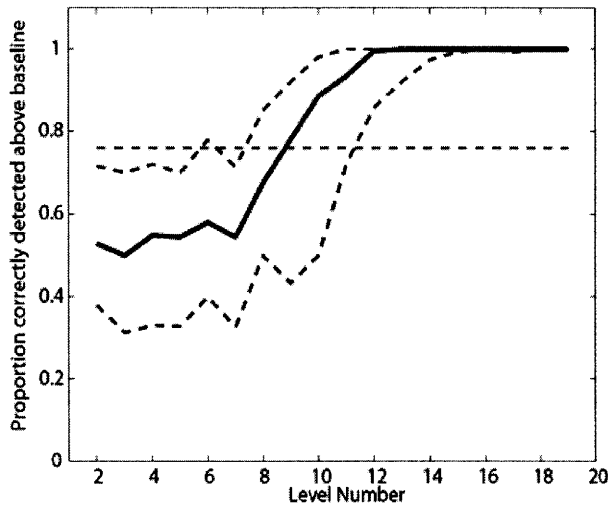
Figure 8: Decoding across all 30 neurons. All models are fit with 25 trials and decoding analysis is performed on the other 15 validation trials. The process is repeated 35 times. See Figure 5 caption for methods. Solid black thick line shows median (50%) performance and dotted lines show 95% confidence bounds. See section 3.7 for detailed description of combining information across multiple neurons.

4.7 Optimal Design of experiments

An important consideration in the design of neurophysiological experiments deals with selecting the appropriate neural recording time, stimulus duration and number of trials for data collection. One would obviously select a recording time large enough to include the portion of the response that is pertinent to encoding the stimulus. Stimulus onset and duration are selected in conjunction with recording time, such that the recorded neurons have sufficiently recovered from encoding the previous stimulus before being presented with another stimulus. Because our inferences are designed to be invariant to the modeled response duration so long as the pertinent response characteristics are captured, we leave the selection of recording time, stimulus onset, and stimulus duration to the experimenter. However, the question of trial number is one that warrants further examination in this framework.

In order to assess to what extent the threshold inference based on multiple trials and levels depends on the number of trials at each level, we used cross-validation to estimate the SS-GLM model using one half of the data (20 random trials) and fit the model with the other half of the data (remaining 20 trials). We compared the MTML threshold estimates derived using half the data to those using the complete dataset for each response. Figure 9 shows that the threshold estimates based on half the data are approximately the same thresholds estimated on

50

the full dataset. However, a minimum number of trials for each stimulation level are needed to accurately estimate the joint distribution of the responses such that the model converges to the ML estimate. The minimum number of trials is dependent on the number of neural spikes in a response, such that neurons with a higher firing rate will tend to require fewer trials to accurately model. This finding has implications for use of smaller number of trials to estimate thresholds accurately, which could dramatically reduce the time needed to record from a neuron.
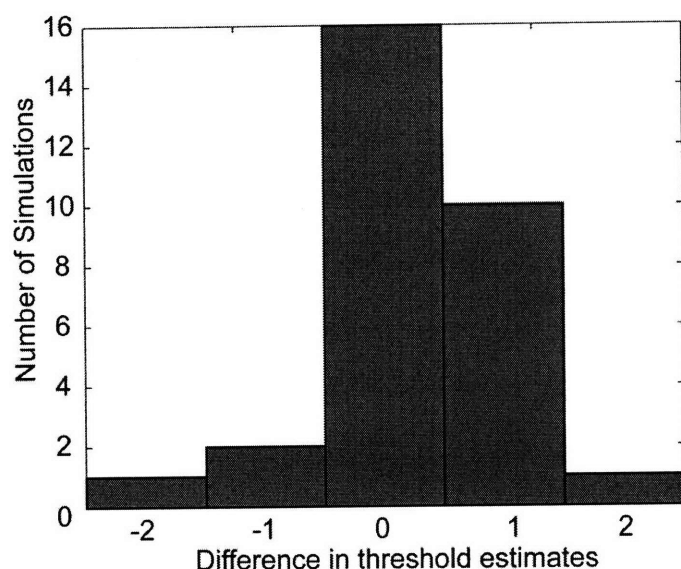
Figure 9: Histogram of differences in threshold estimates between the state-space generalized linear model using the full dataset (40 trials) and that using half of the dataset (20 random trials). Positive differences indicate a smaller threshold estimate using 40 trials and negative numbers indicate smaller threshold estimate using 20 trials. The estimates for the latter are derived by fitting the data on half of the trials and estimating threshold using the other half of the trials.

To study how the number of response trials affects the MTML threshold inference for different response types in more detail, we conduct a theoretical study. While response types greatly differ throughout the auditory system, we focus this theoretical analysis on the response types prevalent in this dataset. We select two responses from the dataset that differ in response characteristics, one with a high background firing rate and a long response duration (Figure 10A1) and one with a lower background firing rate and a short response duration (Figure 10A2). For each response, we fit the SS-GLM model and obtain the ML estimates. Next, we simulate $k = [10, 15, 20, 25, 30, 35, 40]$ new trials based on our encoding parameter estimates for each of the 30 simulation experiments, and run the SS-GLM model on the simulated data and to infer threshold.
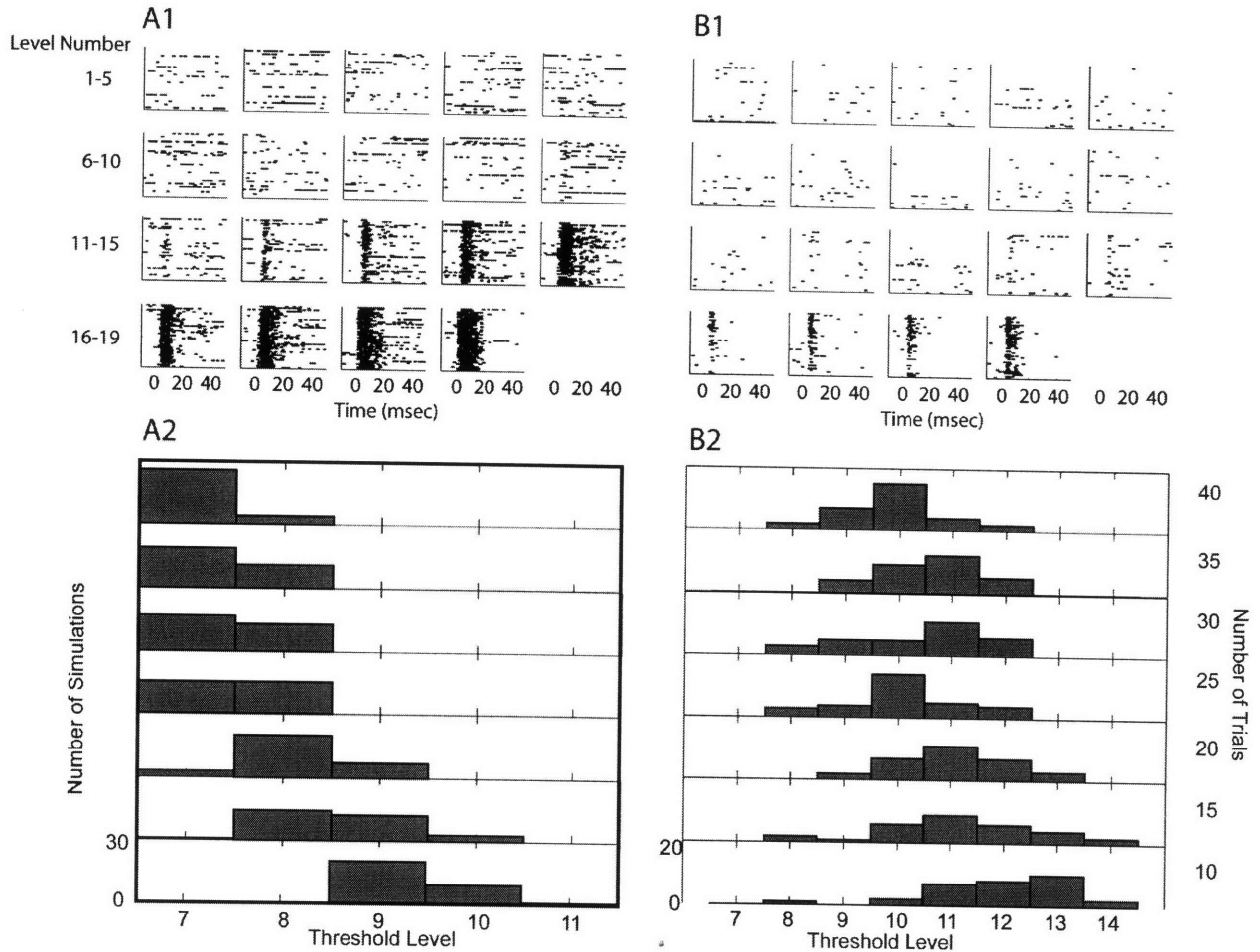
Figure 10: A1: Raster plots for a multi-unit clusters with high firing rate (same data as Figure 1A) A2: Threshold level inferences from simulated data based on maximum likelihood estimates of responses for seven different numbers of trials each with 30 simulations. Number of trials is shown in B2. Data is simulated from maximum likelihood coefficients fit to the data in A1. B1: Raster plots for a maximum likelihood cluster with sparse firing. B2: Threshold level inferences computed similar to A2.

Tighter bounds on threshold estimates are seen for simulated data with a high background faring rate and long response durations, even for as few as $k = 10$ response trials. However, for $k = [25, 30, 35, 40]$, the threshold estimates are restricted to levels 7 and 8, with the propensity to estimate threshold at level 7 increasing with a larger number of trials. For trial numbers $k = 20$ and below, the distribution of threshold inferences increases toward larger threshold levels (Figure 10A). For this response type, the number of trials could have been reduced to 25 or 30 trials without significantly degrading the threshold estimate. This simulation shows a slight bias in the model threshold estimate, which was estimated to be level 8 based on

40 trials of data, and level 7 for the majority of the 30 simulations with 40 trials. The bias is most likely a result of a short modeled duration of stimulus history due to the short recording of baseline activity for the data (see Figure 2)

For response types for which the qualitative difference between the background firing rate and the induced response is smaller, a broader distribution is seen in the threshold level inference (Figure 10B1). Because the background rate is low, different instantiations of the simulated neural response from the ML parameters created a broader range of possibilities for the modeled threshold inference. For $k = [25, 30, 35, 40]$ trials, the threshold inferences are restricted between levels 8 and 12 with a mean at level 10 or 11, while the distribution begins to skew toward higher levels as the number of trials is further reduced. For this neuron, the number of trials could also be reduced to 25 or 30 without significant degradation in the threshold estimate.

Our theoretical simulation results indicate that if the encoding model and MTML inference is of interest, the number of trials for this experiment could be reduced to 25 or 30, representing at least a 25% savings in data collection time. However, these theoretical simulation results also call for caution in interpreting threshold inferences for responses that resemble those in Figure 10A2. For these responses, the distribution of threshold estimates is broader and the threshold inference will therefore be estimated with a lower level of confidence for a smaller number of trials as compared to a response where the ratio of response to background is higher (Figure 10B1). We expect that the certainty with which one could estimate threshold for these neurons would increase with a greater number of trials (>40) and more closely resemble the types of results seen in Figure 10A1.

5 DISCUSSION

5.1 Summary

Although many methodologies for threshold estimation currently exist, the word *threshold* is used to describe the level at which a difference between the baseline and a stimulus-induced response can be detected in one of two conditions: 1) using spike train data from many trials and levels; and 2) using a single spike train each of two levels. These two definitions falling under the same system property of *threshold* lead to very different estimates because the information available in making the first inference is much greater than that available for the second inference.

We propose a new framework that unifies these definitions of threshold while providing more precise estimates of each. Our framework is based on the state-space generalized linear modeling approach of neural modeling (Czanner et al., 2008; Smith et al., 2004). Our analysis of 30 MUs shows that our methods can detect threshold typically 6-8 dB lower than the RLF and SDASA methods, respectively, which estimate threshold using multiple trials at multiple levels. Computing discrimination performance based on single trials allows us to infer threshold lower (typically by 2 dB) than SDT analysis based on only the firing rate. Additionally, the underlying fit of our model is superior to those of currently used models based on firing rate alone.

5.2 Advantages and rationale for SS-GLM model of neural responses

Unlike previous efforts to estimate the response threshold, our inferences are a byproduct, but not the ultimate goal, of our modeling effort. Whereas other estimation methods develop underlying models to derive the estimate, we develop our model to primarily estimate the joint probability of the responses, a byproduct of which is the ability to infer threshold. Therefore, the usefulness of the underlying model extends beyond the inferences mentioned here. The model can serve as a statistical description of the response and from this standpoint, can illustrate the modeled response and its confidence intervals across stimulation level, can provide rate-level functions and confidence intervals or can allow for any other inferences based

on the neural response, such as latency, saturation and dynamic range.

Our methods fill the need of a well-fitting underlying model of neural responses across trial and level. Unlike current estimation methods, the threshold inference made using the SS-GLM model allows for a probabilistic and intuitive interpretation of threshold. Using our methods, the threshold is the response at which one can be reasonably certain that the response is different from baseline. Some of the competing current methods do not have this probabilistic interpretation of threshold that is based on a parsimonious statistical model.

The assessment of the fit of our modeling effort (Figure 2A) attests to this SS-GLM model class providing a good fit for responses of cortical neurons to electrical stimulation. In comparison, the underlying model fits of the rate-based models are very poor. For a rate-based method that assumes a constant firing rate at each level to estimate threshold (homogeneous Poisson model), this model requires estimating one parameter for each level, or 19 parameters for this dataset. Although many current models assume more elaborate neural response descriptions than homogeneous Poisson, such as adding dead time or nonhomogeneous Poisson statistics (Siebert 1968, Winslow and Sachs, 1988; Johnson et al., 1986; Turcott et al., 1994; Johnson, 1996; Heinz et al., 2001), the current method of threshold estimation assumes a homogeneous Poisson process without history for the temporally-restricted response at each level and therefore should be considered as the underlying model for the threshold inference.

In contrast, as we showed in Figure 2A, the fit of the SS-GLM model is highly adequate in capturing the model dynamics, although could be improved with a longer recording duration allowing for a longer modeled history dependence. In addition, the number of parameters used to describe the SS-GLM model is only an order of magnitude larger than the rate-based model. Using an example of 4 ms of history and a 60 ms response, the SS-GLM model captures 56 time bins per level, at 40 trials and 19 levels, for a total of 42,560 data points. The number of estimated parameters here is 56 hidden parameters describing the stimulus effect, 56 initial values of the hidden parameters, 56 variances of the hidden parameters and 3 history

55

parameters, for a total of 171 parameters. The ratio of the number of parameters to data points is less than 0.5%. Keep in mind that in our models, we used 1 ms time bins to describe the stimulus effect. For longer response durations, the binning of the stimulus effect could be made larger such that the number of parameters would not increase dramatically with a longer response duration.

An advantage of our methodology over rate-based methods is that our threshold estimate is independent of the duration of the modeled response. If the response beyond the temporally-restricted response is not significantly different from the baseline response, this extension is ignored in the threshold estimate because the confidence bounds in that region overlap with those at baseline. If the recorded response were longer, the model could theoretically be extended indefinitely and the threshold estimate would remain valid. On the contrary, rate-based methodology typically calls for estimating the response for only a short duration after the stimulus onset (for example, 28 ms for our dataset). Including an offset response and spontaneous activity in the rate-based methods tends to degrade the threshold estimate. For a neuron that shows a propensity for increased spiking with increased stimulus level, increasing the duration of the modeled response would generally decrease the mean spiking rate and increase the variance of the spike rate, thus degrading the estimate. The SDASA method is particularly sensitive to duration increases because both the mean and the standard deviation changes influence the threshold level, while the RLF and rate-based SDT methods are somewhat less sensitive because of the steepness of the neurometric function in the area of the threshold inference.

Three methodological differences exist between our model framework and current frameworks. The first methodological difference is the data itself. By using point process methods, the raw spike trains serve as the inputs to our models, where the data is not aggregated over time or trial. In contrast, rate-based methods assume a homogeneous Poisson model of the data, and therefore estimate the rate of the process by aggregating data over trials

56

and within a time window following the stimulus onset.

Second, the estimation approach differs between the two methodologies, although both are conducted in the likelihood-based framework. Signal detection theory and SDASA methods compute hypothesis tests, a series of likelihood ratios that compare the distribution of responses at each level to the baseline distribution based on an acceptance criterion. In fact, application of the acceptance criterion to a number of independent observations requires compensation for a false positive result, which can occur as the probability of observing data that satisfies the acceptance criterion tends to occur by chance alone. Compensating for the multiple comparisons problem is rarely done by current methods. Instead, our methodology obviates the need for multiple hypothesis tests because we estimate the complete joint density of the responses, and not only the distribution of the response for the null hypothesis as do other methods. We construct confidence bounds and define thresholds as a function of the confidence bounds to avoid the hypothesis testing regime.

The third methodological difference between the two methodologies is the metric used to estimate threshold. When comparing our threshold estimates to those computed using rate-based SDT, we conduct the decoding in the SDT framework by computing the discrimination performance via ROC AUC based on the metric of ML level of the spike train rather than the current metric of the number of spikes in the spike train. We find that our metric allows for a 2 dB improvement in threshold estimates over the rate-based metric in the SDT framework. However, our framework is much broader than this comparison suggests. The basic metric used in our framework for discrimination is the probability that the stimulus-induced response exceeds the baseline response based on a wide variety of data (single trials, multiple trials, multiple levels) as shown in Figure 4A. Although summary statistics can be estimated from these likelihood functions, such as the percent discrimination shown in Figure 4B or the discrimination performance estimated in the SDT framework based on the decoded ML level of the spike train, these summary statistics serve to conform our methods to current methods.

5.3 Extension of threshold discrimination analysis across multiple neurons

In this work, we have illustrated a straightforward framework for combining across responses of different neurons to estimate threshold inferences. Working in the likelihood-based framework greatly facilitates this endeavor because it allows us to conduct decoding analysis simply by maximizing the combined likelihood across multiple units. In this way, the multiple neuron decoding problem is a direct extension of the single neuron decoding problem.

In our discrimination analysis across multiple units (Figure 7 and Figure 8), we observed that discrimination performance is influenced by two factors: the number of neurons or trials over which the decoding is performed and the individual performance of the neurons. In the analysis across multiple classes of neurons (Figure 7) as well as across all 30 MUs in the dataset, we observed that when the number of trials or neurons is held constant, the group performance is similar to that of the best performing neuron. This observation has been noted in behavioral performance studies that have correlated behavioral performance with the performance of the best neuron, according to the lower envelope principle (Parker and Newsome, 1998). Cases in which this holds true is direction discrimination in middle temporal area MT (Parker and Newsome, 1998), flutter vibration discrimination in somatosensory cortex (Romo and Salinas, 2003), in songbird song discrimination in field L avian analogue to the primary auditory cortex (Narayan, 2006) and in tone burst discrimination in the auditory nerve (Siebert, 1965). Our modeling analysis here provides a statistical framework that could serve as an explanation for the lower envelope principle seen in the correlation between neural and behavioral experiments.

5.4. Extension of the model with state-dependent variance

In our current model, the covariance matrix, Σ, of the of the random walk process in (2.9) is constant with level. If one were to imagine that the covariance structure would vary with level, the model could be extended such that

$$\theta_{v,r} = \theta_{v-1,r} + \epsilon_v \tag{2.18}$$

for levels $v = 1, \ldots, V$ where $\epsilon_v = [\epsilon_{v,1}, \ldots, \epsilon_{v,R}]$ is an R-dimensional Gaussian random variable with 0 mean and an unknown covariance matrix Σ_v. Smith et al. (2004) suggest that a state space model in this formulation could be extended such that the level-dependent covariance matrix takes the form $\Sigma_v = \Sigma * g(\theta_{v-1})$, where that the covariance matrix varies as a function of the hidden vector θ_{v-1}. The level-dependent extension to the model adds a good deal of machinery to the EM algorithm. Smith et al. (2004) find that although the confidence bounds using the state-dependent variance model decrease, the estimates of the observation model and inferences are not significantly different from the original model. They argue that if the model fit, as shown by the Kolmogorov-Smirnov goodness of fit test, is adequate for the non-level dependent model as well, that the simpler model is effective at characterizing the data. In our case, the goodness-of-fit is adequate, although could be improved with a longer duration of baseline activity, allowing for model with a longer history dependence, and we therefore believe that our simpler model should be sufficient in describing the data and providing an accurate characterization of threshold.

5.5   Implications for experimental design

Our theoretical analysis showed how to estimate the number of response trials needed to optimize the MTML threshold inference. For the responses seen in this dataset, we estimated that the number of trials could have been reduced by at least 25% (from 40 to 30 or 25) without losing significant confidence in the threshold inference. However, we also showed that the distribution of threshold inferences has a larger spread if the induced response to background response ratio is low. Experimental design is not usually augmented to record different numbers of trials for responses with a low stimulus-induced to background response ratio, although studies have been designed to record different numbers of trials based on measured rates for each stimulus (i.e. Eatock et al., 1991). Our results suggest that if the

number of trials for responses with a low ratio are not increased, those threshold estimates will exhibit greater variability. If a similar certainty of threshold inferences is desired for low and high ratio responses, the number of trials should be increased for these low-ratio responses.

Another experimental design question for neurophysiological studies often concerns the inter-trial duration for experiments that assume no trial dependence. We can use this model class to explore whether a CIF incorporating dependence of trial would be better suited to describe the data. This question could be approached in a parameter exploration session prior to conducting the full experiment, where for a level above the hypothesized threshold, the duration between two trials is varied using the SS-GLM model fit to the responses as described in Czanner et al. (2008). The duration for which the model that does not incorporate trial dependence produces an equally good fit to the model with trial dependence should be chosen for the duration of the full experiment.

REFERENCES

Beitel RE, Snyder RL Schreiner CE, Raggio MW, Leake PA (2000) Electrical cochlear stimulation of deaf cat: comparisons between psychophysical and central auditory neuronal thresholds. *J Neurophysiol* 83: 2145-2162.

Bierer JA, Middlebrooks JC (2002) Auditory cortical images of cochlear-implant stimuli: dependence on electrode configuration. *J Neurophysiol* 87: 478-492.

Box GEP, Jenkins GM, Reinsel GC (1994) *Time series analysis, forecasting and control* 3rd ed. Englewood Cliffs, NJ: Prentice-Hall.

Borisyuk GN, Borisyuk RM, Kirillov AB, Kovalenko EI, Kryukov VI (1985). A new statistical method for identifying interconnections between neuronal network elements. *Biol Cybern* 52: 301-306.

Brillinger DR (1988a) Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol Cyber* 59: 189-200.

Brillinger DR (1992) Nerve cell spike train data analysis: A progression of technique. *J Am Stat Assoc* 87: 260-271.

Britten KH, Shadlen MN, Newsome WT, Movshon A (1992) The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci* 12: 4745-4765.

Brown EN (2005) Theory of Point Processes for Neural Systems In: Chow CC, Gutkin B, Hansel D, Meunier C, Dalibard J, eds. Methods and Models in Neurophysics. Paris, Elsevier, Chapter 14: 691-726.

Brown EN, Nguyen DP, Frank LM, Wilson MA, Solo V (2001) Analysis of neural receptive field plasticity by point process adaptive filtering. *Proc Nat Acad Sci* 98: 12261-66.

Brown EN, Barbieri R, Eden UT, Frank LM (2003) Likelihood methods for neural data analysis. In: Feng J, ed. Computational Neuroscience: A Comprehensive Approach. London: CRC, Chapter 9: 253-286.

Brown EN, Barbieri R, Ventura V, Kass RE, Frank LM (2002) Time-Rescaling theorem and its application to neural spike train data analysis. *Neural Comput* 14:325-346.

Brown EN, Frank LM, Tang D, Quirk MC, Wilson MA (1998) A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J Neurosci* 18: 7411-7425.

Cappé O, Moulines E, Rydén T (2005) *Inference in hidden Markov models*. New York: Springer.

Chernoboy ES, Schramm LP, Karr AF (1988). Maximum likelihood identification of neuronal point process systems. *Biol Cybern* 59: 265-275.

Colburn HS (1981) Intensity perception: Relations of intensity discrimination to auditory-nerve firing patterns. Internal Memorandum, Res. Lab. Electron. MIT, Cambridge, MA.

Czanner G, Eden UT, Wirth S, Yanike M, Suzuki W, Brown EN (2008) Analysis of within-and between-trial neural spiking dynamics. *J Neurophysiol* In Press.

deJong P, MacKinnon MJ (1998) Covariances for smoothed estimates in state space models. *Biometrika* 75: 601-602.

Delgutte B (1986) Peripheral auditory processing of speech information: Implications from a physiological study of intensity discrimination. NATO Advanced Research workshop on the Psychophysics of Speech perception. Utrecht University, The Netherlands.

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J Roy Statist Soc B* 39: 1-38.

Daley D, Vere-Jones D (2003) *An Introduction to the Theory of Point Process.* 2nd ed., Springer-Verlag, New York.

Eatock RA, Weiss TF, Otto KL (1991) Dependence of discharge rate on sound pressure level in cochlear nerve fibers of the alligator lizard: implications for cochlear mechanisms. *J Neurophysiol* 65: 1580-1597.

Eden UT, Frank LM, Barbieri R, Solo V, Brown EN (2004) Dynamic analysis of neural encoding by point process adaptive filtering. *Neural Comp* 16: 971-998.

Fahrmeir L, Tutz G (2001) *Multivariate statistical modeling based on generalized linear model.* 2nd ed., New York: Springer-Verlag.

Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognit Lett* 27: 861-874.

Gaumond PR, Molnar CE, Kim DO (1982) Stimulus and recovery dependence of cat cochlear nerve fiber spike discharge probability. *J Neurophys* 48: 856-873.

Ghosh JK, Delampady M, Samanta T (2006). *An introduction to Bayesian analysis: theory and methods.* New York: Springer.

Green DM, Swets JA (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Greenberg RJ (1998). Analysis of electrical stimulation of the vertebrate retina – work towards a retinal prosthesis. PhD Dissertation, The Johns Hopkins University Baltimore, MD.

Heinz MG, Colburn HS, Carney LH (2001) Evaluating Auditory Performance Limits: I. One-Parameter Discrimination Using a Computational Model for the Auditory Nerve. *Neural Comput* 13: 2273–2316.

Johnson A, Kotz S (1970) *Distributions in Statistics: Continuous Univariate Distributions.* New York: Wiley.

Johnson DH, Swami A. (1983). The transmission of signals by auditory-nerve fiber discharge patterns. *J Acoust Soc Am* 74, 493-501.

Johnson DH, Tsuchitani C, Linebarger, DA, Johnson MJ (1986) Application of a point process model to responses of cat lateral superior olive units to ipsilateral tones. *Hear Res* 21: 135-159.

Johnson DH (1996) Point process models of single-neuron discharges. *J Comp Neurosci* 3: 275-299.

Kass RE, Ventura VA (2001) Spike train probability model. *Neural Comput* 13: 1713-1720.
Kitagawa G, Gersh W (1996) *Smoothness priors analysis of time series.* New York: Springer.

Lane CC, Delgutte B (2005) Neural correlates and mechanisms of spatial release from masking: single-unit and population responses in the inferior colliculus. *J Neurophysiol* 94: 1180-1198.

Lim HH, Anderson DJ (2006) Auditory cortical responses to electrical stimulation of the inferior colliculus: Implications for an auditory midbrain implant. *J Neurophysiol* 96: 975-88.

Lim HH, Anderson DJ (2007). Spatially distinct functional output regions within the central nucleus of the inferior colliculus: Implications for an auditory midbrain implant. *J Neurosci* 27: 8733-8743.

McCullagh P, Nelder JA (1989) *Generalized Linear Models* (2nd ed.). Boca Raton, FL: Chapman & Hall.

McLachlan GJ, Krishnan T (1997) The *EM Algorithm and Extensions.* New York: John Wiley & Sons.

Mendel JM (1995) *Lessons in estimation theory for signal processing, communication and control.* Upper Saddle River, NJ: Prentice Hall.

Middlebrooks JC (2004) Effects of cochlear-implant pulse rate and inter-channel timing on channel interactions and thresholds. *J Acoust Soc Am* 116: 452-468.

Miller MI (1985) Algorithms for removing recovery-related distortion from auditory nerve discharge patterns. *J Acoust Soc Am* 77: 1452-1464.

Miller MI, Mark KE (1992) A statistical study of cochlear nerve discharge patterns in response to complex speech stimuli. *J Acoust Soc Am* 92: 202-209.

Narayan R, Gilberto G, Sen K (2006) Distinct time scales in cortical discrimination of natural sounds in songbirds. *J Neurophysiol* 96: 252-258.

Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point process. *J Am Stat Assoc* 83: 9-27.

Oliver D, Taberner AM, Thurm H, Sausbier M, Arntz C., Ruth P, Fakler B, Liberman MC (2006) The role of BK_{ca} Channel in electrical signal encoding in the mammalian auditory periphery. *J Neurosci* 26: 6181-6189.

Pawitan Y (2001) *In All Likelihood: Statistical Modeling and Inference Using Likelihood.* New York: Oxford Univ. Press.

Parker AJ, Newsome WT (1998) Sense and the single neuron: probing the physiology of perception. *Annu Rev Neurosci* 21: 227-277.

Raggio MW, Schreiner CE (1994) Neuronal Responses in cat primary auditory cortex to electrical cochlear stimulation. I. Intensity dependence of firing rate and response latency. *J Neurophysiol* 72(5): 2334-2358.

Ramachandran R, Davis KA, May BJ (1999) Single-unit responses in the inferior colliculus of decerebrate cats I. Classification base don frequency response maps. *J Neurophysiol* 82: 152-163.

Romo R, Salinas E (2003) Flutter discrimination: neural codes, perception, memory and decision making. *Nat Rev Neurosci* 4: 203-218.

Sachs MB, Abbas PJ (1974) Rate versus level functions for auditory-nerve fibers in cats: tone-bust stimuli. *J Acoust Soc Am* 56: 1835-47.

Shackleton TM, Skottun BC, Arnott RH, Palmer AR (2003) Internaural time difference discrimination thresholds for single neurons in the inferior colliculus of guinea pigs. *J Neurosci* 23: 716-724.

Shepherd RK, Javel E (1997) Electrical stimulation of the auditory nerve. I. Correlation of physiological responses with cochlear status. *Hear Res* 108: 112-144.

Shepherd RK, Hatsushika S, Clark GM (1993) Electrical stimulation of the auditory nerve: the effect of electrode position on neural excitation. *Hear Res* 66: 108-20.

Siebert WM (1965) Some implications of the stochastic behavior of primary auditory neurons. *Biol Cybern* 2: 206-215.

Siebert WM (1968) Stimulus transformation in the peripheral auditory system. In PA Kolers & M Eden (Eds.). *Recognizing patterns*. Cambridge, MA: MIT Press.

Smith AC, Brown EN (2003) Estimating a state-space model from point process observations. *Neural Comp* 15: 965-991.

Smith AC, Frank LM, Wirth S, Yanike M, Hu D, Kubota Y, Graybiel AM, Suzuki W, Brown EN (2004) Dynamic analysis of learning in behavioral experiments, *J Neurosci* 15: 965-91.

Sumner CJ, Tucci DL, Shore SE. (2005). Responses of ventral cochlear nucleus neurons to contralateral sound after conductive hearing loss. *J Neurophysiol* 94: 4234-4243.

Syka J, Popelar J, Kvasnak E, Astl J (2000) Response properties of neurons in the central nucleus and external and dorsal cortices of the inferior colliculus in the guinea pig. *Exp Brain Res* 133: 254-2666.

Taberner AM, Liberman MC (2005) Response properties of single auditory nerve fibers in the mouse. *J Neurophysiol* 93: 557-569.

Tanner MA (1991) *Tools for statistic inference*. New York: Springer-Verlag.

Truccolo W, Eden UT, Fellows MR, Donoghue JP and Brown EN (2005) A point process framework for relating neural spiking activity to spiking history, neural ensemble and extrinsic covariate effects. *J Neurophysiol* 93: 1074-1089.

Turcott RG, Lowen SB, Li E, Johnson DH, Tsuchitani C, Teich MC (1994). A nonstrationary Poisson point process describes the sequence of action potentials over long time scales in the lateral-superior-olive auditory neurons. *Biol Cybern* 70: 209-217.

Vollmer M, Beitel RE, Snyder RL (2001) Auditory detection and discrimination in deaf cats: psychophysical and neural thresholds for intracochlear electrical signals. *J Neurophysiol* 86: 2330-2343.

Warchol ME and Dallos P (1990). Neural coding in the chick cochlear nucleus. *J Comp Physiol A* 166: 721-734.

Winslow RL, Sachs MB (1988). Single-tone intensity discrimination based on auditory-nerve rate responses in backgrounds of quiet, noise, and with stimulation of the crossed olivocochlear bundle. *Hear Res* 35: 165-190.

Witte RS, Otto KJ, Williams JC and Kipke DR (1998). Pursuing dynamic reorganization in auditory cortex using chronic, multichannel unit recordings in awake, behaving cats. *Neurocomputing* 26-27: 593-600.

Zacksenhouse M, Johnson DH, Tsuchitani C (1995) Transient effects during the chopping response of LSO neurons. *J Acoust Soc Am* 98: 1410-1422.

Appendix A: EM algorithm

This section closely follows the EM algorithm derivation for estimating CIF across trials (Czanner et al., 2008). The EM algorithm estimates the ML values of the parameters $\Psi = (\gamma, \theta_0, \Sigma)$, which requires us to maximize the expectation of the complete data log likelihood

$$p(\theta, N | \Psi) = p(N | \theta, \Psi) p(\theta | \Psi) \tag{2.19}$$

where $p(N | \theta, \Psi)$ represents the probability mass function of the point process characterized by the conditional intensity function and $p(\theta | \Psi)$ represents the probability mass function of the hidden state process governed by a random walk. The properties of the ML estimator of consistency, asymptotic normality with asymptotic variance equal to the Fisher information matrix and efficiency hold for Hidden Markov Models in general (Cappé et al., 2005). Applying (2.9) , the random walk model of regularized smoothness constraints, to the likelihood of the Poisson CIF, we can rewrite (2.19), the joint probability density of the data and the hidden parameters as:

$$p(\theta, N | \Psi) = \exp\left(\sum_{v=1}^{V} \sum_{b=1}^{B} n_{b,v} \log(\lambda_v (\Delta b | \theta_v, \gamma, H_v) \Delta) - \lambda_v (\Delta b | \theta_v, \gamma, H_v) \right)$$
$$* \left(2\pi^{-\frac{R}{2}} |\Sigma|^{-\frac{1}{2}} \right)^{V} \exp\left(-\frac{1}{2} \sum_{v=1}^{V} (\theta_v - \theta_{v-1})^{T} \Sigma^{-1} (\theta_v - \theta_{v-1}) \right) \tag{2.20}$$

The EM algorithm is an iterative process to compute ML estimates and has two steps. In the first step (E-step), we compute the expectation of the complete data log likelihood given all the data and the estimates of the unobserved parameters and hidden processes (Ghosh et al., 2006):

$$Q(\Psi|\widehat{\Psi}^i) = E\left[\sum_{v=1}^{V}\sum_{b=1}^{B} n_{b,v}\log(\lambda_v(\Delta b|\theta_v,\gamma,H_v)\Delta) - \lambda_v(\Delta b|\theta_v,\gamma,H_v))||N,\widehat{\Psi}^i\right]$$
$$-\frac{1}{2}RVlog(2\pi) - \frac{1}{2}Vlog|\Sigma| \qquad (2.21)$$
$$-\frac{1}{2}E\left[\sum_{v=1}^{V}(\theta_v - \theta_{v-1})^T\Sigma^{-1}(\theta_v - \theta_{v-1})||N,\widehat{\Psi}^i\right]$$

where i represents the iteration of the EM algorithm and the notation $E[\cdot\ ||N,\Psi^i]$ indicates the expectation given the data and the parameter estimates at iteration i. From the hidden state expectation, we need to estimate the following quantities:

$$\theta_{v|V} = E[\theta_v||N,\Psi^i]$$

$$W_{v,v+1|V} = E[\theta_v\theta_{v+1}||N,\Psi^i]$$

$$W_{v|V} = E[\theta_v^2||N,\Psi^i]$$

We compute these quantities in the E-step using the forward filter, smoothing filter and state-space covariance algorithms. In the in the second step (M-step), we maximize the expected value of the data log likelihood, $Q(\Psi|\widehat{\Psi}^i)$, with respect to the estimated parameters, Ψ, at iteration i and obtain $\widehat{\Psi}^i$ such that $\max Q(\Psi|\widehat{\Psi}^i) = Q(\Psi^{i+1}|\widehat{\Psi}^i)$. The essence of the EM algorithm is that the complete data log likelihood, $Q(\Psi|\widehat{\Psi}^q)$, can act as a proxy for the likelihood, $L(\Psi)$, and any increase in $Q(\Psi|\widehat{\Psi}^i)$ leads to an increase in $L(\Psi)$ (Cappé et al., 2005). In general, starting from any point, the EM algorithm can be expected to converge to the local maximum (Ghosh et al., 2006) since the likelihood is non-decreasing, leading to a task of monotonic optimization (Cappé et al., 2005).

E-Step I: Forward filter:

In this first step of the E-step, we use the adaptive filter algorithm to compute the posterior density of the hidden parameters $\theta_{v|V} = E[\theta_v||N,\Psi^i]$ as follows (Eden et al., 2004):

One step
prediction:

$$\theta_{v|v-1} = \theta_{v-1|v-1} \qquad (2.22)$$

One step
prediction
variance:

$$W_{v|v-1} = W_{v-1|v-1} + \Sigma \qquad (2.23)$$

Posterior
variance:

$$W_{v|v} = -\left[W_{v|v-1}^{-1} + \frac{\partial^2 \log\,(\lambda_v(b\Delta|\theta_v,\gamma,H_b))}{\partial\theta_v^T \partial\theta_v} n_{v,b} \right.$$
$$\left. - \lambda_i(b\Delta|\theta_v,\gamma,H_b)\Delta \right] \qquad (2.24)$$

Posterior
mode:

$$\theta_{v|v} = -\left[\theta_{v|v-1}^{-1} \right.$$
$$\left. + W_{v|v} \sum_{b=1}^{B} \lambda_v(b\Delta|\theta_v,\gamma,H_b) \right.^{-1} \frac{\partial\,(\lambda_v(b\Delta|\theta_v,\gamma,H_b))}{\partial\theta_v} * [n_{v,b} \qquad (2.25)$$
$$\left. - \lambda_v(b\Delta|\theta_v,\gamma,H_b)\Delta] \right]\Big|_{\theta_v=\theta_{v-1}}$$

These equations are solved in the order shown with the initial conditions $\theta_{1|0} = \theta_0$ and $W_{0|0} = 0$.

E-Step II: Fixed Interval Smoother: in this step of the E-step, we compute $\theta_{v|V}$ and $W_{v|V}$ given estimates of the posterior mode $\theta_{v|v}$ and variance $W_{v|v}$ from the adaptive filter algorithm (Mendel 1995; Brown et al., 1998)

$$\theta_{v|V} = \theta_{v|v} + A_v(\theta_{v+1|V} - \theta_{v+1|v}) \qquad (2.26)$$

$$A_v = W_{v|v}W_{v+1|v}^{-1} \qquad (2.27)$$

$$W_{v|V} = W_{v|v} + A_v(W_{v+1|V} - W_{v+1|v})A_v^T \qquad (2.28)$$

The algorithm is evaluated for $v = V - 1, ..., 1$ and for initial conditions $\theta_{V|V}$ and $W_{V|V}$ derived from the adaptive filter algorithm.

E-Step III: State space covariance algorithm: this algorithm allows us to estimate the remaining quantity $W_{v,v+1|V} = E[\theta_v\theta_{v+1}||N,\Psi^i]$ needed to evaluate the expectation of the complete data log likelihood as follows (de Jong and Mackinnon, 1988)

68

$$Cov[\theta_v, \theta_u || N, \Psi^i] = A_v Cov[\theta_{v+1}, \theta_u || N, \Psi^i] \tag{2.29}$$

From (2.29) it follows that:

$$W_{v,v+1|V} = Cov(\theta_v, \theta_{v+1}) + \theta_{v|V}^T \theta_{v+1|V} \tag{2.30}$$

M-Step: In the M-step, we maximize over the rth expected value of the complete log likelihood to obtain the next iteration of estimates $\Psi^{i+1} = (\gamma^{i+1}, \theta_0^{i+1}, \Sigma^{i+1})$. However, in order to solve maximize over the complete data log likelihood, we need to be able to evaluate

$$E[\lambda^S(b\Delta|\theta_v) || N, \Psi^i] = E[\exp(\sum_{r=1}^{R} \theta_{r,v} g_r(b\Delta) || N, \Psi^i)]$$

which can be approximated as a Taylor expansion about the rth value in the parameter $\theta_{r,l|L}$ due to the index functions:

$$E[\exp(\sum_{r=1}^{R} \theta_{r,v} g_r(b\Delta) || N, \Psi^i)] = \exp(\theta_{r,v|V}) + \frac{1}{2}\exp(\theta_{r,v|V}) W_{r,r,v|V} \tag{2.31}$$

where $\theta_{r,v|V}$ is the rth element and $W_{r,r,v|V}$ is the rth diagonal element. Since a closed-form solution for maximizing with the respect to the history parameters does not exist (since they are non-stationary), we use Newton-Raphson algorithm (Fahrmeir and Tutz, 2001) to iteratively find

$$\gamma_{m+1}^i = \gamma_m^i - \left[\frac{\partial^2 Q(\gamma_m^i, \theta_0^i, \Sigma^i | \gamma^i, \theta_0^i, \Sigma^i)}{\partial \gamma^2}\right]^{-1} \left[\frac{\partial Q(\gamma_m^i, \theta_0^i, \Sigma^i | \gamma^i, \theta_0^i, \Sigma^i)}{\partial \gamma}\right],$$

where m is the number of iterations of the newton-Raphson algorithm at the ith iteration of the EM algorithm. The algorithm stops when the difference between two consecutive estimates of $|\gamma_{m+1}^i - \gamma_m^i|$ is smaller than 10^{-2} for all J elements in the vector and the value of the history parameter is updated in the EM algorithm ($\gamma^{i+1} = \gamma_{m+1}^i$). There are closed form solutions for maximizing with respect to the other parameters θ_0, Σ:

$$\Sigma^{i+1} = \frac{1}{L}\sum_{l=1}^{L} E[(\theta_v - \theta_{v-1})(\theta_v - \theta_{v-1})^T || N, \Psi^i] \tag{2.32}$$

$$\theta_0^{i+1} = E[\theta_1 || N, \Psi^i] \tag{2.33}$$

The EM algorithm is stopped with the absolute change in the parameters falls below 10^{-2} for all vector components of the parameters of if the value of the complete log-likelihood function does not increase. The EM algorithm is initialized with the GLM estimates of the stimulus and spike history components and $\frac{1}{10}I$, where I is the RxR identity matrix.

Appendix B: Estimation of confidence bounds for stimulus effect

The confidence bounds for the instantaneous firing rate in each bin can be simulated using Monte Carlo. The bounds are computed as follows:

1) Draw a sample size of $m = 1{:}M$ from a Gaussian distribution with mean vector $[\theta_{1,r|v}, \dots, \theta_{V,r|v}]$ and covariance matrix W

2) For every draw m, compute $\lambda_m^s(b\Delta|\theta_{v,r|v})$. Because our model assumes no variation in firing rate across trial, average over trial.

3) For every l, order the values of $\lambda_m^s(b\Delta|\theta_{v,r|v})$ from largest to smallest.

4) The $100(1-a)\%$ confidence bound for $\lambda^s(b\Delta|\theta_{v,r})$ is such that $\frac{a}{2} = \frac{m_1}{M}$ and $1 - \frac{a}{2} = \frac{m_2}{M}$. The confidence bounds at each $b\Delta$ are $[\lambda_{m_1}^s(b\Delta|\theta_{v,r|v}), \lambda_{m_2}^s(b\Delta|\theta_{v,r|v}),]$. In our analysis, we use $M = 300$.

Chapter 3

Modeling auditory psychometric functions and population estimates of proportion data using a state-space model analysis

1.   ABSTRACT

In behavioral studies of auditory perception, the psychometric function is used to model the relationship between the stimulus intensity and a proxy for perceived intensity of sound, such as the proportion of correct stimuli detections or the reaction time needed to perform the task.   We illustrate how a Bayesian analysis state-space model can be used to represent psychophysical functions, mapping stimulus level to one or multiple metrics of behavioral performance. We model the detection performance and reaction time of a guinea pig stimulated with multi-site silicon substrate probes in the inferior colliculus (Pfingst et al., 2007) and a unilaterally deafened primate implanted with a multichannel scala-tympani electrode array in the basal turn of the cochlea (Pfingst et al., 1985). The stimulus consisted of a 250 pulses/second biphasic, charge balanced current pulse train of 200 ms duration and 200 μs/phase pulse width from .01 to 50 μA presented in logarithmic steps at 2 dB increments. Animals were trained to respond with a go/no-go if the stimulus was detected and their reaction time was also recorded (Pfingst et al., 1975a, 1975b). We present these analyses as an alternative to the existing approaches of fitting continuous sigmoidal functions to the behavioral data, and explain how inferences, such as the current level for which behavioral performance reaches some value, can be estimated from the model. We also show that our Bayesian estimate of threshold as the stimulus level for which we are reasonably certain (> 0.95) that the reaction at that level is different from baseline, the width of the range of levels for which the reaction is different from baseline and the maximum performance can be computed easily in our framework. In addition, we illustrate that another measure of behavioral performance, such as reaction time, can be included in the model, such that the joint distribution can be estimated simultaneously for both

variables, and inferences can be made using either or both measurements of behavioral performance. Using the guinea pig and primate datasets, we find that the concurrent estimation of psychometric functions using both proxy variables leads to larger confidence intervals and therefore higher threshold inferences if both proxy variables do not move in concert. The higher threshold estimate may be appropriate when each proxy variable is thought to equally represent the perceived sound intensity. In addition, we examine how this class of stochastic models can be extended to population studies, including responses from multiple subjects within multiple groups, where between-subject differences offer an important source of variance. We illustrate a random effect state-space (RESS) model to estimate the population and individual subject responses, and to perform dynamic inference of performance between the two populations and within the same population, avoiding hypothesis testing. In a study of the neural correlates of learning an auditory association task, we conclude that differences in neural responses with duration of learning are evident at the start and conclusion of the task for the association data, but not for the no-association data. In addition, we find that the differences between the two groups are found in the neural responses at the start of the experiment, which could indicate a neural correlate in the Basal Ganglia of behaviorally-observed auditory association learning. Both Bayesian modeling methods offer a practical approach to the analysis of proportion data within a theoretical framework for auditory experiments.

2. INTRODUCTION

The behavioral psychometric function is a change in some observed parameter, such as stimulus detection, discrimination, or task reaction time, in response to a change in some dimension of the stimulus. For auditory stimuli, that dimension may be intensity, amplitude modulation or frequency; for visual stimuli, color intensity or frequency, for tactile stimuli, pressure or frequency of perturbation. These behavioral measurements are studied to investigate how genetic or physiological differences (Youngentob and Margolis, 1999; Rance et al., 1998; Dougherty et al., 1998), stimulus changes (Dougherty et al., 1998), and electrical stimulation (Miller et al., 2000; Hughes et al., 2001) cause deviations from stimulus detection or discrimination performance seen in normal subjects. Inferring accurate psychometric functions in subjects with a prosthesis and comparing performance to normal subjects is particularly important in the design of improved prosthetics (Friesen et al., 2001; Fu and Shannon, 1998).

To develop a model of the psychometric function that is both specifically tailored toward creating a probabilistic model of behavioral responses and to providing strong methodological ties to neurophysiological methods for threshold detection, we introduce a state-space model of threshold detection. We focus on the problem of behavioral threshold estimation using auditory midbrain implants to complement our recent work on neural threshold detection in a similar experimental model (Chapter 2). Following recently introduced models of learning (Smith et al., 2003; Smith et al., 2004; Smith et al., 2007), we represent the behavioral task responses as a Binomial probability model, and the unobservable change in responses with increased stimulus level as a state equation (Kitagawa and Gersh, 1996). We estimate the model parameters using Bayesian methods to give a precise statistical statement about the stimulus level at which a detectable behavioral response occurs. Incorporating findings that suggest that in addition to detection probability, other metrics of performance, such as reaction time, serve as accurate proxy variables for the internal representation of stimulus intensity, we develop additional models to estimate psychometric functions using more than one proxy variable. In addition, we

74

illustrate a general framework for inference in our discrete dynamic model of the psychometric function with and without the second proxy variable such as reaction time.

Auditory perception studies are often extended from the framework described above by incorporating multiple subjects and multiple groups. Often, multiple groups are needed to account for the large between-subject variation in auditory perception studies, in order to extract features common to each population. However, the between- and within-subject differences are typically not quantified, and the performance is averaged across subjects within a group (e.g. Barnes et al., 2005; Miller et al., 1999; Werner et al., 1992). If the experiments are particularly lengthy, for instance during auditory learning association experiment with animals, some subjects may be withdrawn from the experiment before the experiment concludes. However, current quantification of population performance does not always incorporate the increased uncertainty in the population performance estimates when the subject pool reduces during the duration of the experiment. In addition, simple averages of individual performance to estimate population performance do not capture the inter-subject variability present in different subject groups.

To characterize this extended framework of auditory experiments and estimate both population and individual responses, we develop a random effects model for multiple subjects performing a task within several subgroups. Random effects models are commonly used statistical paradigms to analyze time series measurements (Fahrmeir and Tutz 2001; Laird and Ware, 1982; Stiratelli et al, 1984). Random effects models have been used in medical and biological research and have recently been applied to behavioral learning experiments (Smith et al., 2005).

We illustrate the use of the random effect state-space model by analyzing an experiment exploring the neural correlates underlying the learning of an auditory association task. In this study, two groups of animals ran through a T-maze while listening to a high- and a low-tone stimulus. For the association group, the experiment was designed such that subjects learned to

associate the sound of the stimulus with a right or left turn, and correct association was rewarded to facilitate learning. In contrast, the non-association group was rewarded on random trials, and therefore did not learn a correct turn association. Neural responses from all the animals in both groups were recorded and the percent of activated neurons during several events from each experimental trial served as a potential neural correlate of learning. We analyze these neural proportion data and illustrate dynamic estimates of between- and within-group differences of neural performance. Our methodology has the advantage of practical and computationally efficient implementation within the Bayesian MCMC framework, a characterization of between- and within-subject variation and the representation of auditory learning as a dynamic process.

3.   METHODS

3.1   Behavioral psychometric function estimation and inference

3.1.1. Modeling methodology

Our modeling framework is developed in a state-space framework. State space models with unobservable processes are also called hidden Markov models (HMM) or latent process models (Kitagawa and Gersh, 1996; Fahrmeir and Tutz, 2001; Smith and Brown 2003; Smith et al., 2004). HMMs are Markov chains in noise, with the noise contributing to the unobservable or hidden aspect of the model (Cappé et al., 2005). If the Markov chain is called, X_k, what we observe is a stochastic process, Y_k, that is linked to the Markov chain through the noise. For our task of estimating psychometric functions, the first, state-space, component defines the evolution of the internal representation of stimulus intensity, and thus disambiguates the perceived level of the stimulus from the task of correctly identifying whether a stimulus is present. In other words, the state-space component defines the transition probabilities of the hidden variable , X_k, in the Markov chain. The second, observed, component relates the degree of correct detection of the stimulus, Y_k, to the unobservable representation of stimulus intensity, X_k,. Because the degree of correct detection is thought of as a proxy for the internal representation of stimulus intensity, we define the unobservable representation of stimulus intensity such that increases in the internal representation of stimulus intensity increase the degree of correct detection and decreases in internal representation of stimulus intensity decrease the degree of correct detection. Similar analysis has been used to describe learning in behavioral experiments (Smith and Brown 2003; Smith et al., 2004; Smith et al., 2007).

We formulate our model and estimate the underlying representation of the stimulus intensity state variable at each level after seeing the behavioral responses for all stimulus levels. Therefore, to identify significant deviations in performance, we first estimate the underlying state process for the observed data, and second, we compute the confidence bounds for the individual performance curves and the population by computing the probability that the

77

performance at each stimulus level differs from the performance in the absence of stimulation.

To model the auditory stimulus detection task, we assume that an animal is asked to detect a stimulus over multiple sessions and trials. For a stimulus played at a given level, v ranging from $1, ..., V$, the animal produces n_v go responses indicating that the stimulus is detected from a total of m_v presentations of the stimulus. For each level, p_v governs the probability of detecting the stimulus at that level. The observable probability model defines the probability distribution associated with detecting a stimulus played at a level v. We model this probability distribution as Binomial probability mass function:

$$\Pr(n_v | p_v) = \binom{m_v}{n_v} p_v^{\,n_v} (1 - p_v)^{\,m_v - n_v} \qquad (3.1)$$

The binomial assumption captures the notion that the performance at a given stimulus intensity has the same success probability, ignoring processes such as learning and adaptation. Psychophysicists argue that for well trained subjects, the assumption is justified (Blackwell, 1952), but one should be aware of the assumption when applying the model to novel datasets.

We assume that the detection of the response at each level v is governed by some unobservable process x_v, here, internal representation of the stimulus intensity, and that this process governs the progression of detection as a function of stimulus level. The internal representation of the stimulus intensity x_v is defined as a Gaussian random-walk:

$$x_v = x_{v-1} + \epsilon_v \qquad (3.2)$$

where ϵ_v is a vector of independent Gaussian random variables with mean 0 and variance σ_τ^2. Often, a drift term is appropriate in the model of how the internal representation of the stimulus intensity at one level relates to the internal representation of the stimulus at the consequent level, since it is generally assumed that the psychometric function is typically non-decreasing. The drift reflects the belief that a louder stimulus produces a more accurate response. Following a previous learning study (Smith and Brown, 2003), the internal representation of the stimulus intensity with the drift term can be modeled as:

$$x_v = \alpha + x_{v-1} + \epsilon_v \tag{3.3}$$

where α represents the drift term and is positive in this application.

We relate the internal representation of the stimulus intensity, x_v, to the performance on the detection task, p_v, by defining $\Pr(x_v | p_v) = f(x_v | p_v = P_v)$, where $f(\cdot)$ is a functional constraint that ensures that the probabilities of detection remain between 0 and 1. To model this constraint, we define p_v in terms of x_v with a logistic relationship as

$$p_v = \frac{\exp(x_v)}{1 + \exp(x_v)}, \tag{3.4}$$

for all levels $1, \ldots, V$. Therefore, as the perceived stimulus level x_v increases to infinity, p_v increases to 1 meaning perfect detection, and in the absence of a perceived level, p_v reaches 0 meaning that the stimulus is never detected.

Additionally, sometimes data is available that describes the reaction time r_v pertaining to each stimulus presentation and is related to the internal representation of the stimulus intensity as $\Pr(x_v | r_v) = g(x_v | r_v = R_v)$. Studies (i.e. Pfingst et al., 1975a, 1975b) suggest that reaction time and number of correct responses are both proxy measurements for the internal representation of the stimulus intensity. When reaction time data is available, we expand the model to simultaneously model both the relationship between psychophysical performance and level, as well as reaction time and level. We model the relationship between p_v and x_v as in (3.4) and model the relationship between reaction time r_v and x_v as:

$$r_v = b + ax_v + z_v \tag{3.5}$$

where a and b are random variables and z_v is a vector of independent Gaussian random variables with mean 0 and variance σ_z^2. Another possibility for the relationship between r_v and x_v is that of a lognormal distribution, since the reaction times are strictly positive:

$$\log(r_v) = b + ax_v + z_v. \tag{3.6}$$

However, in this thesis, we illustrate our methodology with the reaction times using the Gaussian distribution as shown in (3.5).

3.1.2. Bayesian implementation of model

For models with hidden parameters or missing data, maximum likelihood estimates of parameters, θ, and hidden processes, x, can also be computed using a Bayesian approach as an alternative to the EM methods of estimation (as shown in Chapter 2 of the thesis). Bayes formula weighs prior beliefs about the true parameter values with the congruency of those beliefs with the observed data (Congdon, 2003). The goal of the Bayesian approach is to estimate the posterior probability density of x and θ. Although the true generating parameters x^{TRUE} and θ^{TRUE} are hidden and unknown, respectively, our observation data reduces the uncertainty of the value. This goal comes out of Bayes formula:

$$p(\theta, x | N) = \frac{p(\theta)p(x|\theta)p(N|x,\theta)}{p(N)} \qquad (3.7)$$

where $p(\theta)$ is the prior probability density for θ, $p(x|\theta)$ is the joint probability density of the hidden processes x, and $p(N|x,\theta)$ is the likelihood of the data, N, given the parameters and hidden process (Ghosh et al., 2006). These distributions describe our beliefs about the uncertainty inherent in the parameter values. Typically, the denominator does not participate in the estimation as it is just a normalizing constant, such that posterior probability density sums to unity.

Often, and in the two model classes that will be described here, the posterior, $p(\theta, x | N)$, cannot be solved analytically or numerical integration is not feasible, however, probability distributions for the likelihood and the prior are computationally realistic. Bayesian techniques such as Markov Chain Monte Carlo methods circumvent the problem of solving for the posterior by drawing samples from the unnormalized posterior distribution, $p_u(\theta, x | N) = p(\theta)p(x|\theta)p(N|x,\theta)$. MCMC methods are specifically designed for our problem specification where the joint posterior distribution of the data is specified in terms of several conditional and marginal distributions, such that the joint posterior is difficult to calculate, but the conditional posterior at different hierarchical levels is easier to compute. At the point of

convergence, MCMC procedures guarantee that the random draw will be from the target joint posterior distribution. These samples can then be used to obtain asymptotic distributions of the posterior. The method relies on the Law of Large Numbers that guarantees that our sampling estimates of the sample statistics will be representative of the population with a sufficiently large number of samples (Gamerman, 1997; Ghosh et al., 2006).

The inference step involves estimating the posterior distribution $p(\theta, x | N)$ that incorporates all our prior knowledge of the parameters with data observations. In this problem of modeling the auditory psychometric function, we want to estimate the unknown parameters $\theta = (x_0, \sigma_\tau^2)$ for the basic model. When a drift parameter is included in the model, the unknowns become $\theta = (x_0, \sigma_\tau^2, \alpha)$, and when reaction time is also included, the unknowns expand to $\theta = (x_0, \sigma_\tau^2, \sigma_z^2, \alpha, a, b)$. The joint probability density of the internal representation of the stimulus intensity state given by Eq (3.2) is

$$p(x|\theta) = \prod_{l=1}^{L} \frac{1}{\sqrt{(2\pi\sigma_\tau^2)}} \exp\left\{-\frac{1}{2}\sigma_\tau^{-2}(x_v - x_{v-1})^2\right\} \qquad (3.8)$$

and the joint probability of the data from Eq (3.1) is

$$p(N_{1:V}|x, \theta) = \prod_{d=1}^{D} \binom{n_v}{m_v}(p_v)^{n_v}(1 - p_v)^{m_v - n_v}. \qquad (3.9)$$

With a broad choice of the prior $p(\theta)$, the posterior density will greatly resemble the complete data likelihood, since using likelihood methods, we estimate $p(x|\theta)p(N|x, \theta)$ while using the Bayesian methodology, we estimate $p(\theta)p(x|\theta)p(N|x, \theta)$ if the prior, $p(\theta)$, is sufficiently broad (Ghosh et al., 2007). However, Bayesian methods have the advantage of not requiring rewriting of the estimation algorithms with each change of the model because by specifying the distributions of all priors, sampling methods are available to draw samples from the posterior distributions.

We model the variance of the error term of the random-walk σ_τ^2 as a gamma function,

parameterized by $\alpha_{\sigma_\tau^2}$ and $\beta_{\sigma_\tau^2}$ such that it has wide support, because the gamma is the conjugate prior of the Gaussian distribution (Gammerman, 1997; Congdon, 2003; Ghosh et al., 2006). We model the potential drift term as a random variable with a uniform distribution $[\alpha_1, \alpha_2]$ where α_1 and α_2 are greater than zero, reflecting the belief that the performance increases with loudness. We will illustrate fits of the psychometric function with and without the drift term. Finally, we model the parameters a and b describing the transformation from the internal representation of the stimulus intensity to the reaction time by uniform probability distributions $[-a_1, a_1]$ and $[-b_1, b_1]$ with wide support and the variance of the error term for reaction time σ_z^2 also as a gamma function parameterized by $\alpha_{\sigma_z^2}$ and $\beta_{\sigma_z^2}$.

We select the Bayesian framework for this modeling effort to easily change the model using different functional forms for $f(\bullet)$ and $g(\bullet)$ because our implementation allows us to simulate samples from the posterior distributions. We use MCMC methods to compute the posterior probability by simulating stationary Markov Chains (Gilks et al., 1996; Congdon, 2003). We conduct the MCMC simulations in the WinBUGS (Spiegelhalter et al., 1996-2004; Lunn et al., 2000) environment and the Matbugs (Murphy and Mahdaviani, 2005) interface for MATLAB (The Mathworks, Natick, MA). After we specify the priors and the joint probability of the data, WinBUGS simulates the posterior probability density using Gibbs sampling techniques (Geman and Geman, 1984; Gelfand and Smith, 1990; Gilks et al., 1996) which decomposes the parameter space by sampling from the posterior distributions of each parameter individually, and is thus particularly suitable for problems of high dimension (Gamerman, 1997; Kosh, 2007).

To assess convergence of the MCMC simulation, we examine the mixing of the three Monte Carlo chains simulated in the analysis and computed the Brooks-Gelman-Rubin statistic to compare the between- and within-chain variance (Gelman and Rubin 1992) such that it is <1.2 for all parameters (Kass et al. 1998; Smith et al., 2005). For this analysis, we employ 10,000 Monte Carlo iterations per chain (in addition to 1,000 burn-in iterations) and achieve

convergence in ≪ 1 minute of CPU time (for the models with and without reaction time) on a Pentium 4M laptop computer.

3.1.3 Estimating behavioral detection and reaction time

The previous section stated the model description and the implementation used to obtain point estimates of the model parameters. These estimates are selected to minimize an error criterion with respect to the posterior distribution. In this work, we present the point estimates taken as the medians of the posterior distributions because the median minimizes the expected absolute error otherwise known as the absolute deviation loss (Ghosh et al., 2006; Kosh, 2007). Other point estimates could include the mean, minimizing the expected square error or the maximum-a-posteriori estimate that minimizes a loss function which is equal to 0 when the parameter estimate equals the true parameter and 1 otherwise (Ghosh et al., 2006; Kosh, 2007). We refer to the median estimates of the probability of detection as

$$\hat{p}_v = \frac{\exp{(\hat{x}_v)}}{1 + \exp{(\hat{x}_v)}}. \qquad (3.10)$$

and to the median estimate of the reaction time as:

$$\hat{r}_v = \hat{b} + \hat{a}\hat{x}_v + \hat{z}_v. \qquad (3.11)$$

where the circumflex (hat) denotes the estimates are taken as the median (50% percentile) values of the Monte Carlo estimates of the parameters.

3.1.4. Calculation of Bayesian credible intervals

We use summary statistics from the set of Markov samples to estimate the posterior credible intervals. The credible intervals convey the size of the parameter distributions. Unlike frequentist techniques used in Chapter 2, the concept of credible intervals conveys a subtly different notion compared to frequentist confidence intervals. In Bayesian statistics, a credible interval of $100(1 - q)\%$ is the interval in which the true parameter values lie $100(1 - q)\%$ of the time. Because the parameters are random variables in the Bayesian framework, the credible intervals are computed from samples of the estimated posterior distribution. However, in

frequentist statistics, the parameters are viewed as unknown and deterministic, not random variables. Therefore, the $100(1 - q)\%$ confidence bounds are those in which the true value of the parameter is found with a probability $100(1 - q)\%$ (Ghosh et al., 2006; Koch, 2007). Although for arbitrary priors, the confidence and credible intervals are not necessarily similar, for objective priors, frequency interpretations are approximately true using the normal approximation to the posterior distribution (Ghosh et al., 2006). Because the frequentist confidence bounds cannot be computed analytically, asymptotic approximations are used or bootstrap techniques (Efron and Tibshirani, 1993) to create artificial datasets many times to derive the confidence bounds (see Chapter 2). In the Bayesian framework, the samples are generated from the posterior, and the credible intervals can be simply found from the MCMC samples. Therefore, in our Bayesian framework here, the $100(1 - q)\%$ credible intervals can be estimated by taking the $\frac{q}{2}$ and $1 - \frac{q}{2}$ quantiles of the sample density.

3.1.5. Description of experimental data

To illustrate our methodology of psychometric function estimation and inference, we model the detection performance and reaction time of a guinea pig stimulated with multi-site silicon substrate probes, each with 16 sites, in the inferior colliculus (Pfingst et al., 2007) and a unilaterally deafened primate implanted with a multichannel scala-tympani electrode array in the basal turn of the cochlea (Pfingst et al., 1985). The stimulus consisted of a 250 pulses/second biphasic, charge balanced current pulse train of 200 ms duration and 200 μs/phase pulse width from .01 to 50 μA presented in logarithmic steps at 2 dB increments. Animals were trained to respond with a go/no-go if the stimulus was detected and their reaction time was also recorded (Pfingst et al., 1975a, 1975b). Animals were trained to depress a button mounted on the cage floor until a stimulus was presented. Following the stimulus presentation, the animals had to release the button quickly in order to receive a food reward. Early releases triggered a time out. Reaction times for each response were also recorded. Typically, each stimulus condition was

repeated 20 times in a session and each session was repeated 15 times. Stimuli were presented in random order. No learning is expected since animals are well trained with the paradigm and the false alarm rate was maintained below 20% or the animal was retrained (Miller et al., 2000).

3.2. Random effects state space model of population of proportion data

3.2.1. Random effects state space model overview

As in the model of the psychometric function, we assume a state space formulation for the analysis of the proportion of activated neurons in the auditory learning experiment. A state equation, tracking the temporal evolution of the underlying auditory learning process, and an observation equation, relating the observable data and the unobservable learning process for each event, define the state-space framework. We formulate a random effects state-space (RESS) model because we would like to characterize and compare changes in response properties across two groups and for individual subjects within each group. To do this, we assume that the individual subjects are random members of the probability distribution characterizing the population, with the population learning process as the mean. The model proposed here has roots in the analysis of population learning developed by Smith et al. (2005).

The data we observe includes the number of activated neurons and the total number of neurons for each animal in each group during each event and during each stage (see section 3.2.6 below). The events are the times during which tasks are performed by the animal while running through specific locations in the T-maze. These numbers of activated neurons are calculated according to the procedure defined in Barnes et al. (2005) for both groups.

As in the analysis of auditory psychometric functions, we formulate our model to estimate the learning state at each day (stage) based on the neural responses for all the days for all the subjects. Therefore, to identify when significant deviations in learning activity occur, we first estimate the underlying learning state process and second, we compute the confidence

bounds for the individual learning curves and the population by computing the probability that the activity during each stage after the first differs from the response on the first day with no learning.

The RESS model was chosen to model the proportion of activated neurons across times for animals in the two groups for several reasons. First, there are multiple animals in each group and each animal is seen as a sample of a general population. Therefore, the variability between the animals in a single group can be described using a random effect component in the model. Second, the underlying motivation between comparing the neural responses between the two groups is to examine the learning of an association that should have occurred for the one group and should not have occurred for the other. We model this underlying learning process that slowly varies with time as a hidden process using a state space formulation. Lastly, a model is needed to estimate the full distributions of responses over time for the two groups. Because some days had fewer data, knowing the full distributions and the inferred confidence bounds will create a more accurate model of the data and the significance of any estimated changes across stage than simple averages across animals.

To define the RESS model, we assume that we have computed the number of activated neurons for each stage (time) for both groups according to the Barnes et al. (2005) criteria defined previously for each of A_g animals ranging from $a_g = 1, .., A_g$ for each group g and each animal perfoms tasks or events, e, whose number ranges from $e = 1, ..., E$. Time is defined as stages, d, of learning, where d ranges from $d = 1, ..., D$. To define the observation equation, we let $n_{d,e,g}^a$ define the number of activated neurons, $m_{d,e,g}^a$ define the number of total neurons and $p_{d,e,g}^a$ define the probability of a correct response for animal a in group g during stage d for event e. Further, we assume that the dynamics of the probability of correct response are governed by an unobservable learning process $x_{d,e,g}$ for each event and group that defines the dynamics of the learning as a function of stage. The relationship between the observed data

and the percentage of activated neurons can thus be expressed as a binomial probability mass function as follows:

$$\Pr\left(n_{d,e,g}^a \,\middle|\, p_{d,e,g}^a, x_{d,e}\right) = \binom{n_{d,e,g}^a}{m_{d,e,g}^a} (p_{d,e,g}^a)^{n_{d,e,g}^a} (1 - p_{d,e,g}^a)^{m_{d,e,g}^a - n_{d,e,g}^a}. \qquad (3.12)$$

where $p_{d,e,g}^a$ is governed by the logistic function:

$$p_{d,e,g}^a = \frac{\exp\left(\mu_g + \beta_{e,g}^a x_{d,e,g}\right)}{1 + \exp\left(\mu_g + \beta_{e,g}^a x_{d,e,g}\right)}. \qquad (3.13)$$

Notice that we assume that in the absence of learning, naïve animals perform at chance level, governed by the parameter μ_g. We assume that all naïve animals within a group g perform at the same level. When learning of an auditory association does occur for animal a in group g during for event e, the parameter $\beta_{e,g}^a$ modulates the extent of the learning. These two parameters offer extensions of the model of the psychometric function presented previously. In the psychometric function analysis, we assumed that $\beta_{e,g}^a = 1$ and $\mu_g = 0$, while in this model, these parameters are estimated. Each animal-specific random effect component is independently drawn from the population distribution defined by

$$\beta_{e,g}^a \sim N\left(\beta_{e,g}, \sigma_{e,g}^2 I_{A_g x A_g}\right).. \qquad (3.14)$$

where $\beta_{e,g}$ describes the population mean and $\sigma_{e,g}^2 I_{A_g x A_g}$ describes the population variance, and $I_{A_g x A_g}$ is the identity matrix of dimension of the number of animals in each group A_g. The proportion of activated neurons in the population follows the same logistic regression as the individual animal estimates:

$$p_{d,e,g} = \frac{\exp\left(\mu_g + \beta_{e,g} x_{d,e,g}\right)}{1 + \exp\left(\mu_g + \beta_{e,g} x_{d,e,g}\right)}. \qquad (3.15)$$

We model the unobservable learning state that governs learning from one day (stage) to another day (stage) as a random walk:

$$x_{d,e,g} = x_{d-1,e,g} + \epsilon_{d,e,g}, \qquad (3.16)$$

where the random walk components are independent, Gaussian and zero-mean variables,

$\epsilon_{d,e,g} \sim N(0, \sigma^2_{\epsilon_{e,g}})$.

3.2.2. Bayesian analysis of neural responses using an RESS model

The objective of the analysis is to estimate $\theta = (\beta^0_{e,g}, x_{0,e,g}, \sigma^2_{\epsilon_{e,g}} \sigma^2_0)$, where $\beta^0_{e,g} = \{\beta^1_{e,g}, \dots, \beta^A_{e,g}\}$, and from these estimates to calculate the probability of activated neurons $p^a_{d,e,g}$ for each animal a in group g during event e, as well as the probability of activated neurons for the entire population over each group, $p_{d,e,g}$. As in the models of psychometric functions, we use Bayesian sampling to estimate the unobservable and unknown parameters. Again, this Bayesian sampling technique is an alternative to the Expectation Maximization algorithm (as used in Chapter 2 of the thesis) for performing maximum-likelihood estimation with an unobservable process or missing observations.

The goal of the Bayesian analysis is to compute the posterior probability of θ and x, defined by Bayes' rule as:

$$p(\theta, \beta, x | N) = \frac{p(\theta)p(x, \beta | \theta)p(N | x, \beta, \theta)}{p(N)}, \qquad (3.17)$$

where $p(\theta)$ is the prior probability density for θ and $p(x|\theta)$ is the joint probability density for the learning state process, which is defined (2.17) as:

$$p(x_{e,g}, \beta_{e,g} | \theta) \qquad (3.18)$$

$$= \prod_{d=1}^{D} \frac{1}{\sqrt{(2\pi\sigma^2_{\epsilon_{e,g}})}} \exp\left\{-\frac{1}{2}\sigma^{-2}_{\epsilon_{e,g}}(x_d - x_{d-1})^2\right\} \prod_{a=1}^{A} \frac{1}{\sqrt{(2\pi\sigma^2_{\epsilon_{e,g}})}} \exp\left\{-\frac{1}{2}\sigma^{-2}_{\beta_{e,g}}(\beta^a_{e,g} - \beta^0_{e,g})^2\right\}.$$

Notice that the joint probability density described for the RESS model is an extension of the psychometric model because the psychometric model does not include the state-modulation parameters. The joint probability density of the data is defined as:

$$p(N_{1:D,g,e} | x, \beta, \theta) = \prod_{d=1}^{D} \prod_{a=1}^{A} \binom{n^a_{d,e,g}}{m^a_{d,e,g}} (p^a_{d,e,g})^{n^a_{d,e,g}} (1 - p^a_{d,e,g})^{m^a_{d,e,g} - n^a_{d,e,g}}. \qquad (3.19)$$

The prior probability density, $p(\theta)$, for the variance of the error term, $\sigma^2_{\epsilon_{e,g}}$, the mean of the population modulation parameter, $\mu_{\beta_{e,g}}$, and the variance of the animal modulation parameters, $\sigma^2_{\beta_{e,g}}$, are modeled as gamma probability densities with parameters $\alpha_{\sigma^2_{\epsilon_{e,g}}}$ and $B_{\sigma^2_{\epsilon_{e,g}}}$, $\alpha_{\mu_{\beta_{e,g}}}$ and $B_{\mu_{\beta_{e,g}}}$, and $\alpha_{\sigma^2_{\beta_{e,g}}}$ and $B_{\sigma^2_{\beta_{e,g}}}$ respectively.

To perform the Bayesian estimation, we again use Monte Carlo Markov Chain (MCMC) methods (Congdon 2003; Gilks et al. 1996) using winBUGS simulation environment (Lunn et al. 2000; Spiegelhalter et al. 2004) and a Matlab (The Math Works, Natick, MA) interface using Matbugs (Murphy and Mahdaviani, 2005) as discussed in the psychometric estimation section. Similarly to the psychometric function methods, the Bayesian credibility intervals of $100(1 - \alpha)\%$ are computed directly by taking the $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ quantiles of the Monte Carlo sample probability density. Convergence is assessed using the Brooks-Gelman-Rubin statistics and the stationarity of the chains, as previously discussed. For these models and using the data described below, we employ 10,000 Monte Carlo iterations per chains (in addition to 1,000 burn-in iterations) to achieve convergence in <25 minutes CPU time on a Pentium 4-M laptop computer.

3.2.3. Estimating the individual learning

The maximum likelihood estimates of the individual probability of activated neurons for animal a in group g during for event e is defined as

$$\hat{p}^a_{d,e,g} = \frac{\exp(\hat{\mu}_g + \hat{\beta}^a_{e,g}\hat{x}_{d,e,g})}{1 + \exp(\hat{\mu}_g + \hat{\beta}^a_{e,g}\hat{x}_{d,e,g})}, \tag{3.20}$$

where the circumflex (hat) denotes the estimates are taken as the median (50% percentile) values of the Monte Carlo estimates of the parameters. Similarly, the population probability of activated neurons estimate in group g during for event e is defined from Eq. (2.9) as:

$$\hat{p}_{d,e,g} = \frac{\exp(\hat{\mu}_g + \hat{\beta}_{e,g}\hat{x}_{d,e,g})}{1 + \exp(\hat{\mu}_g + \hat{\beta}_{e,g}\hat{x}_{d,e,g})}. \tag{3.21}$$

It is important to point out that this approach does not simply average the individual animal probability of activated neurons curves. Instead, we simultaneously estimate the population and individual learning curves within the Bayesian framework. Such is possible because each modulation parameter is a random effect for each subject and has the same probability density for each day of the experiment. The probability density of the underlying learning state changes from day to day and depends on the previous learning state.

3.2.4. Within- and between-group differences in neural responses

The objectives of this framework are twofold. First, assess whether the neural responses, as measured by the percentage of activated neurons, change as learning occurs within each group and within each event. Second, determine whether the neural responses differ between the two groups within each event. Both inferences can be carried out easily in this framework given our estimates of the probability distribution associated with the estimates of the probability of activated neurons. By comparing the credible intervals for each curve, we can compute the probability that one curve is greater than the other to estimate our certainty of differences within a group and between two groups.

To assess whether within-group differences exist, we look for a stage for which the neural response of activated neurons for the untrained animal (stage 1) is significantly different from another stage within a group and an event. Given a level of certainty of .90, the learning stages are stages for which $\Pr(\hat{p}_{d,e,g} > \hat{p}_{1,e,g}) \geq 0.90$, for stage $2 < d \leq D$ for group g during for event e. We can assess whether the probability exceeds our criterion level of 0.90 by examining whether the confidence bounds for the first stage and consequent stages overlap. Likewise, to assess whether between group differences exist for any stage for a single event, we look for a stage or groups of stages for which $\Pr(\hat{p}_{d,e,g_1} > \hat{p}_{d,e,g_2}) \geq 0.90$ by looking for stages for which the 90% confidence bounds do not overlap.

3.2.5. Model cross-validation

We cross-validate the model using posterior predictive checks (PPC) (Gelman et al., 1995; Congdon, 2003). This method creates new data, Z, based on the model specifications, Φ, and assesses whether the new data reproduces important aspects of the actual data, Y,. Bias and precision do not concern this cross-validation technique because if the model is appropriate, the new data should have these features. Critiques of this technique argue that it is a conservative test (Bayarri and Berger, 2000).

To compute the PPC measure, one must decide on a criterion, $C(Y; \Phi)$, emphasizing important data features such as the ratio of observed variance to mean or skewness measure (Congdon, 2003). In this analysis, we select the ratio of observed variance to mean of the original and simulated data as the criterion. The same criterion is computed from the newly generated data, $C(Z; \Phi)$. At each iteration of the cross-validation, we obtain $C(Z^i; \Phi^i)$ for the new data and $C(Y; \Phi^i)$ for the original data and compute the proportion of iterations for which $C(Z^i; \Phi^i)$ exceeds $C(Y; \Phi^i)$ as:

$$P_{PPC} = \sum_{i=1}^{I} \frac{C(Z^i; \Phi^i) > C(Y; \Phi^i)}{I}.$$

(3.22)

Values near 0 or 1 (in practice, above .9 and below .1 are used) indicate that the new data differs from the original data, and values close to .5 indicate a close relationship between the new and original data based on the checking function (Congdon, 2003). We compute the PPC statistic for every iteration of the model (10,000 Monte Carlo iterations, with an additional 1000 burn-in samples).

3.2.6. Description of experimental data

3.2.6.1.   Experimental paradigm: behavioral training

Neural spike activity was chronically recorded in the sensorimotor striatum from two groups of rats performing training on a T-maze task with implanted headstage assemblies. The first group of rats, the association group, was trained on a conditional reward-based task for 24-

63 daily sessions. The second group of rats, the non-association group, was trained for 30 daily sessions. Recording began on the first day that the animals received training and continued through acquisition training (stages 1-5) and overtraining (stages 6-15). The non-association group was trained for stages 1-13. All animals performed approximately 40 trials per day, separated by 1-3 minute inter-trial intervals. Behavioral performance for each animal during each trial was also recorded.

The association group of animals learned to run down a maze and turn right or left as indicated by the auditory stimulus cue sounded prior to the turn in order to receive a reward. During each trial, the rats heard an auditory click (70 dB SPL) followed by the gate opening. The rat proceeded to run down the long arm of the T-maze and half-way down the maze, one of two tones sounded and remained on until the rat reached one of the goals (1 kHz or 8 kHz tones, 80 dB SPL). The association group received a reward when the animal correctly identified one of the tones with the correct goal. For the non-association group, both arms of the goal contained a reward for rewarded trials and neither contained a reward for non-rewarded trials. Tone-goal arm assignments were randomized and counterbalanced for both groups of rats. The results of the association group from a previous experiment (Barnes et al., 2005) determined the ratio of trials containing rewards. Therefore, the difference between the two groups of animals was that learning an association occurred for the seven association group animals but not for the eight non-association group animals.

3.2.6.2.   Data acquisition and analysis: behavioral data

The behavioral performance of each animal for each training day in the association group was measured by the accuracy of the responses averaged over all the trials during the session. In order to combine the neural responses from different animals in the same group, the following stages of learning were used by the Graybiel lab do define the training, based on the individual animal's behavioral performance in associating the pitch of the tone with the direction of the turn:

- stage 1 – first training session

- stage 2 – second training session

- stage 3 – first training session with > 60% correct

- stage 4 – first training session with > 70% correct

- subsequent stages – subsequent pairs of consecutive sessions with >72.5% correct performance

This staging procedure defines the relationship between behavioral performance and neural data as well as combines neural data across multiple subjects.   For the non-association animals, the data was also staged to make them comparable to the associative animals. The following staging categories were used for the non-associative group:

- stage 1 – first training session

- stage 2 – second training session

- stage 3 – first session that animals received 60% or more of the reward (day 5 of training)

- stage 4 – first training session with > 70% of the reward (day 11 of training)

- Stage 5 – first of two training days when rats received 72.5% or more of the reward (days 13 and 14).

- subsequent stages (6-13) – subsequent pairs of consecutive sessions with >72.5% of the trials containing a reward. Stage 6 were training days 15 and 16, stage 7 were training days 17 and 18 and so on until days 29 and 30 corresponding to stage 13.

3.2.6.3.   Data acquisition and analysis: neural recordings

Headstages carrying tetrodes with seven independently moveable microdrives (six for recording and one for reference) were mounted on the scull of male Sprague-Dawley rats as described in Barnes et al. (2005). Tetrodes were gradually lowered through the brain toward

the striatum during the one week period following the surgery. Unit activity recorded on each

tetrode and the activity was sorted into single units and units were classified into putative

projection neurons, putative fast-firing interneurons and putative tonically firing interneurons

based on their activity during a pre-trial 500-ms baseline period. Only the putative projection

neurons (medium spiny neurons) were analyzed in this study. The neurons were classified as

task or non-task related by the peri-event time histograms (PETH) for each unit for each time-

stamped task event (warning cue, gate opening, locomotion onset, tone onset, turn onset and

offset and goal reaching). Task related responses were defined as responses in which the

spike counts in four or more consecutive 20-ms bins, with at least one of the bins occurring in

± 200 ms peri-event time windows, had 2 or more spikes and exceeded the criterion level, let as

two standard deviations above the mean activity recorded during the pre-trial baseline 1900 to

1400 ms before the warning cue (Barnes et al., 2005). For those units that did not fire during

the pre-trial baseline, the Graybiel lab defined task-related responses as epochs with four or

more consecutive bins with spike counts of at least 2.

4. RESULTS

4.1. Analysis of psychometric functions using Bayesian sampling

To illustrate the use of stochastic models to estimate psychometric functions, we use data from Bryan Pfingst and colleagues of behavioral responses to electrical stimulation. Two sets of responses are available, one from a guinea pig and one from a primate. Both responses are from a single subject and are shown in Figure 11A and B, where the proportion of correct trials, an average of the number of indications that the animals detected the stimulus, is shown in gray asterisks. The two sets of data from the two different animals differ in several ways. First, the discrimination accuracy of the guinea pig never reaches the accuracy level of the primate. At his best performance, the guinea pig detects just over 80% of trials containing the stimulus (Figure 11A), while the primate consistently performs at 95-100% accuracy (Figure 11B) over a 20 µAmp range of stimulus levels (10-30 µAmp). A second difference in performance concerns the steepness of the increase in correct detection to reach asymptotic performance. The guinea pig begins to show improved performance slightly below 10 µAmp and does not reach high levels of performance of approximately 75% correct until 20 µAmp. The primate's performance increases more steeply over a smaller range of current levels. The primate's performance begins to increase at 5 µAmp and reaches the animal's peak performance level of approximately 95% at 10 µAmp, a 5 µAmp range compared to more than a 10 µAmp range for the guinea pig. However, one similarity that both animals exhibit is decreased performance at higher stimulus levels. The guinea pig performance reaches a maximum at 30 µAmp and degrades by approximately 10% until 50 µAmp for the last 3 current levels tested. The primate's performance remains steady until 30 µAmp, and decreases to approximately 65% over the last 2 current levels tested.

95
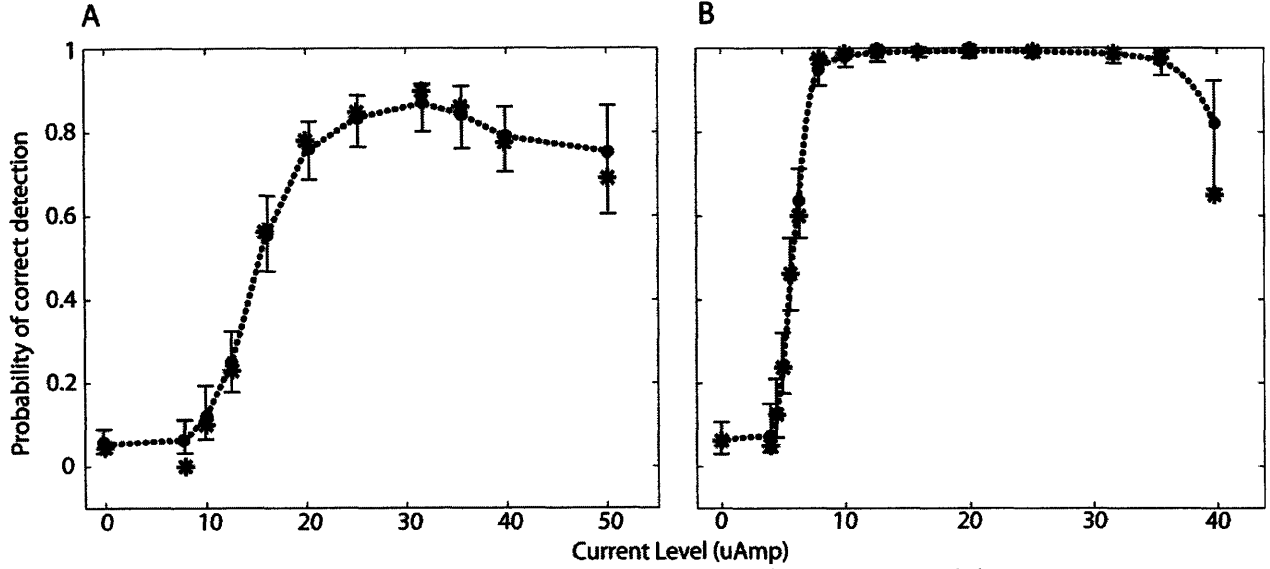
Figure 11: Model with no drift component for guinea pig (A) and primate (B) responses to train of brief current pulses. Responses are averaged over multiple trials for one subject only in A and B. Data is shown in gray asterisks, median model estimates are shown as black circles and error bars span the 95% credible intervals.

4.1.1. Bayesian model of psychometric function with one dependent variable

To model these two responses, we conducted a Bayesian analysis using an MCMC implementation for two sets of models, one with a drift term and one without. We first show the results of the model without a drift component. The model without a drift term assumes that the underlying intensity perception follows a random walk given by $x_v = x_{v-1} + \epsilon_v$, for $l = 1, ..., L$, where $\epsilon_v \sim N(0, \sigma_\tau^{-2})$ with $x_0 = \log\left(\frac{p_s}{1-p_s}\right)$, where p_s is the false alarm rate, the initial probability of a correct response with the stimulus is off. We fix the initial value of x_0 because we assume that the probability of a correct response when the stimulus is off is just at our statistical level of significance. We use gamma priors for σ_τ^2 to ensure that the variance of the noise is always positive. The gamma prior is chosen to have a wide range such that the mean is equal to 1 and variance to 0.2. The probability of correct detection is computed using Equation (2.17) . This model is fit at the sampling rate of the data.

The probability of correct detection along with the 95% credible intervals are shown in black in Figure 11A and B. Although the model largely follows the shape of the data, important differences exist between the model fits to the guinea pig and primate data. Because the

guinea pig data varies slowly from level to level and the performance at extreme current levels is not steady, the model tends to deviate more from the data. The deviation arises from the random walk term, which tends to induce the memory of the last data point into the current point estimation. The deviations of data and model are most obvious at the second data point and the last data point (highest level), since the data deviations from the previous point are largest there. The smoothing of the random walk term in the model leads the model to underestimate the degree of these data movements. An excellent example of this phenomenon in the primate dataset is at the highest level, where the actual performance drops almost 35%. However, the modeled performance drops only about 20% because of the memory from the previous measurement. This inherent memory property of the model is particularly appropriate to estimates of the psychometric function since large fluctuations in the data are not expected and transitions are expected to be smooth (see Discussion).

The sizes of the estimated confidence bounds also reflect the amount of fluctuation in the data from point to point. Since the movements from level to level in the data tend to be large in the guinea pig data compared to the primate data, the credible intervals are also larger for the guinea pig data. The exception in the primate data are the error bars for the model estimates during the rapid increase in performance from 5 µAmp to 10 µAmp, where the error bars are on par with the guinea pig data due to the large movements in performance with each data point.

In learning experiments, scientists sometimes assume that the animal has an underlying learning bias. The probability of a correct response from the animal is therefore a function of not only the stimulus and some inherent noise in the system (accounting for lapses in attention or motor execution), but also of a bias to answer in the positive or the negative. In the studies examined in this work, on the other hand, the animals are well trained, having attained an acceptably high level of correct responses and a low level of false alarms, so bias issues should not account for performance. However, for the type of data examined here, experimenters and

modelers follow the assumption that the performance should be relatively monotonic with level (i.e. Wichmann, 2001a; 2001b; Treutwein and Strasburger, 1999). We can incorporate this assumption into our model by adding a drift term to the random walk equation. This method is used in learning experiments to incorporate bias. The random walk equation is now altered to

$x_v = \alpha + x_{v-1} + \epsilon_v$, for $v = 1, ..., V$, where $\epsilon_v \sim N(0, \sigma_\tau^{-2})$ with $x_0 = \log(\frac{p_s}{1-p_s})$, where p_s again the false alarm rate which we set to .05. Again, gamma priors for σ_τ^2 are used and the probability of correct detection again is computed using Equation (3.10) Now, the internal representation of stimulus intensity is a function of that at the previous level, the noise and a drift term, which in this case is positive and would lead to an increase in performance over an unbiased random walk model. To ensure that the drift term is positive, we use a uniform prior for α, with a range between .2 and 2.

We implemented the model with the additional drift term on both the guinea pig and the primate data and the original data along with the model results are shown in Figure 12. First, examine the guinea pig data and model. Unlike the model with no drift (Figure 11A), the last modeled point is now virtually at the same level as the best detection performance at 30 and 35 μAmp. Although the model is not monotonic (because of the 40 μAmp point), it exhibits closer monotonicity than the model without drift. In fact, because the credible intervals overlap, statistical tests at the 95% confidence level will not distinguish between the modeled animal performance from 25 μAmp until 50 μAmp, indicating that the animal has achieved asymptotic performance. Notice that the drift model offers the greatest compensation from non-monotonic performance only at the points where the data is the most non-monotonic. At points where the data exhibits monotonicity, the drift and no drift models yield essentially the same model estimates (compare to Figure 11A).
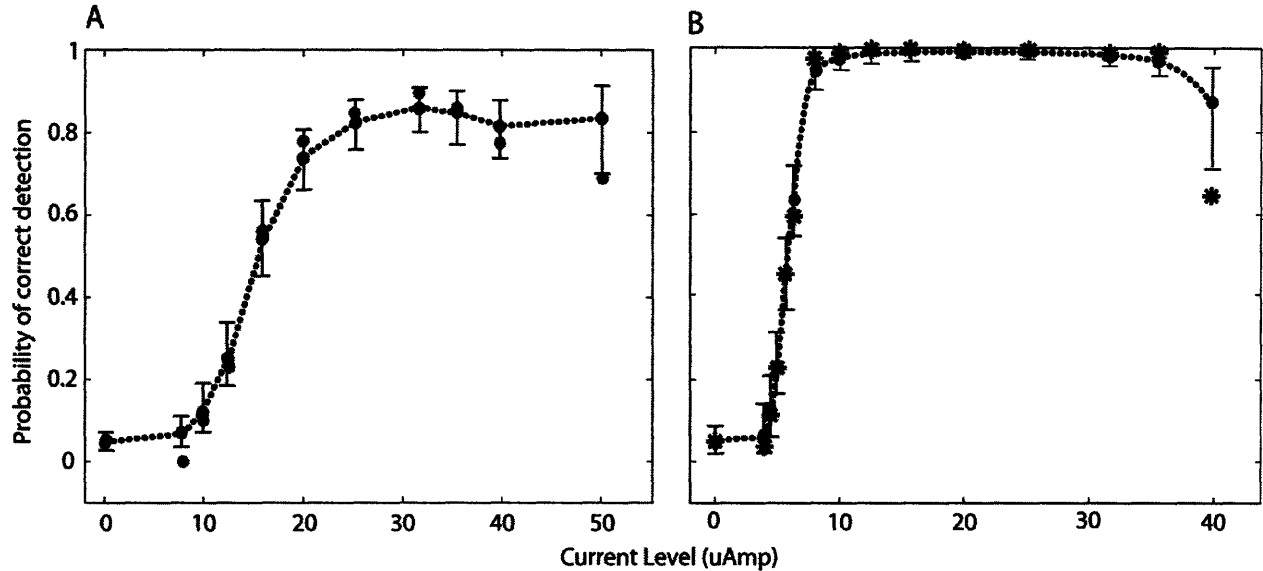
Figure 12: Model with additional drift component for guinea pig (A) and primate (B) in response to train of brief current pulses. Symbol description same as Figure 11.

The differences between the drift and no drift models seen in the guinea pig data are also seen in the primate data (Figure 12B). Again, the fit of the two models is very close for data that does not exhibit monotonicity (all but the performance at the two highest stimulus levels). However, the performance at the highest level is so poor compared to the asymptotic performance seen between approximately 10 to 35 μAmp that even the additional drift term cannot create a monotonic model fit. However, compared to the no drift model, the drift model conveys more monotonicity in the data at the last point. This dataset suggests that even with generously wide priors on the size of the drift term, such a large lapse in performance at the high stimulus level cannot lead to a monotonic model of the data (see discussion).

The stochastic models present an inherently different approach than the continuous, monotonic models. One difference, as previously illustrated in Figure 11 and Figure 12, and discussed in the discussion, is the lack of assumption of monotonicity in the psychometric function, which allows one to test the hypothesis of monotonicity given model assumptions. The second difference, which we illustrate here, is that the model is discrete. However, the discrete nature of the model does not preclude model estimates at fine level steps.

We illustrate model performance estimates here at level steps 10 times finer than the sampling of the data. The model sampling here is chosen for illustrative purposes only and can be as fine as needed. In fact, the sampling frequency should be selected by the experimenter to coincide with that necessary to achieve appropriately smooth psychometric functions suitable for the project. The drift and no-drift models at the finer sampling rates for both the guinea pig and primate data are shown in Figure 13. In addition, these figures replicate the model estimates at the sampling frequency of the data shown in Figure 11 and Figure 12. Two main observations can be made from comparing the model estimates and confidence bounds at the two different sampling rates. First, the model estimates at the actual data points using the finer and coarser estimates are essentially identical. These observations can be explained by the sampling of the stochastic differential equations executed by the MCMC algorithm. The sampling, which in the limit approximates the stochastic differential equation solution to the random walk (and random walk with drift) model, depends on the initial conditions at the data points which remain the same, independent of the sampling rate. Second, the confidence bounds of the finely-sampled model are identical to those of the coarser model at the data points, but bow out slightly in between data points, with the largest bowing out occurring between the two data points. Because the intermediate points in the finely-sampled model are computed using the same random walk patterns (with or without drift), the estimates of the intermediate points are heavily based on the model estimates at the data points. Since the credible intervals scale with the model sampling, the credible intervals should be approximately steady with the sampling rate, in that the degree of bowing is not a function of the model sampling rate (Smith et al. 2003).
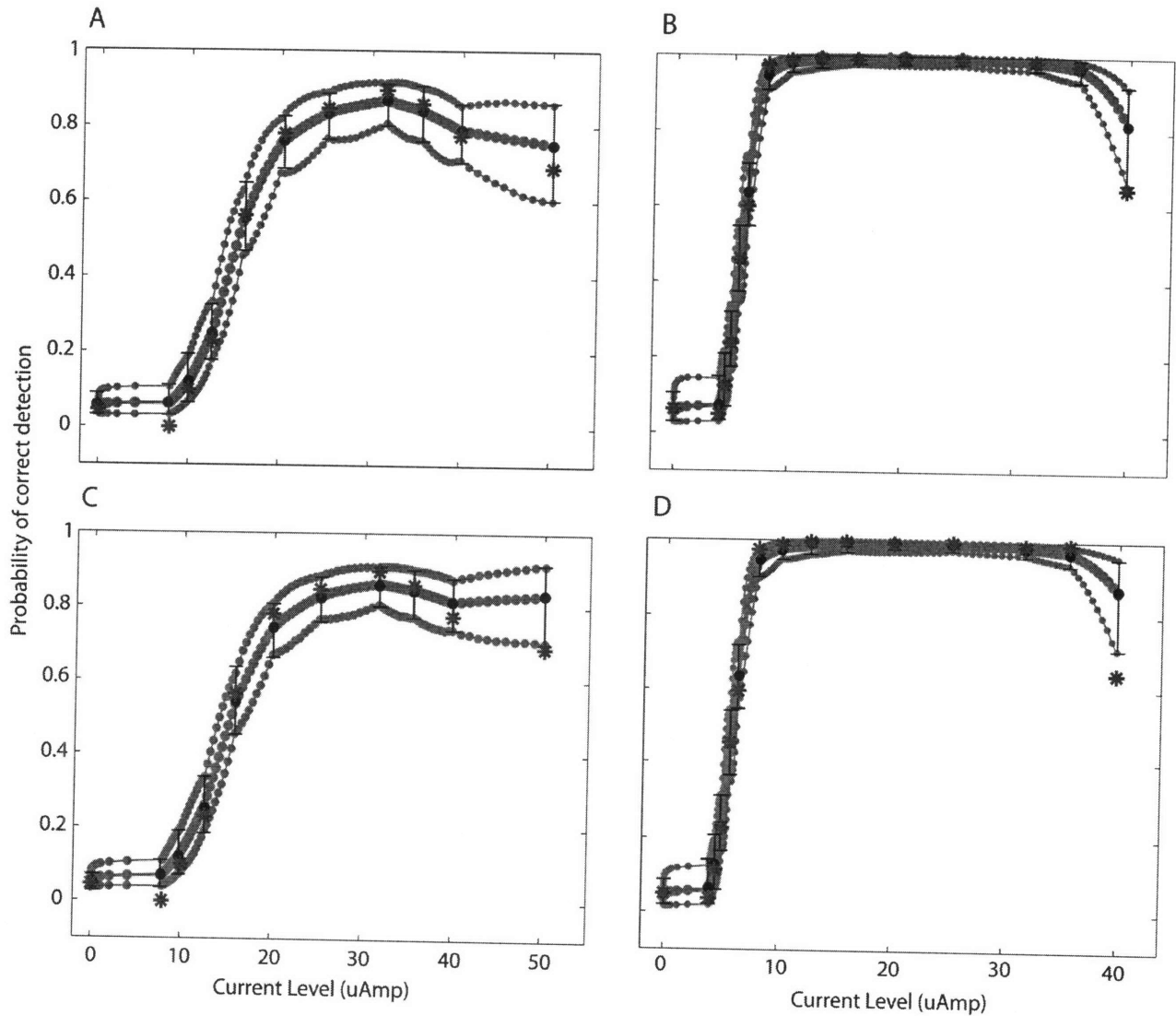
Figure 13: Model estimates at finer intervals than data points. Models without the additional drift term are shown in the top row (A and B), while the models including an additional drift term are shown in the bottom row (C and D). Guinea pig responses shown in left panels (A and C) and primate responses shown in right panels (B and D). Dark gray circles show the median and 95% credible interval estimates of model sampled at 10 times the data sampling, black circles show the median model estimates at the sampling rate of the data and black error bars show the 95% credible intervals for the model sampled at the rate of the data.

4.1.2. Inference within Bayesian framework

Inferences, such as threshold and psychometric function width and slope can easily be found in our stochastic framework. We illustrate how to find four inferences from the primate and guinea pig data, based on the drift and no drift models: the Bayesian threshold, the width and slope of the psychometric function and the point at which performance reaches a specific criterion (50% correct). Inferential statistics are simple and intuitive in this Bayesian framework

101

since the estimation process involves drawing samples from the joint posterior model

distribution.
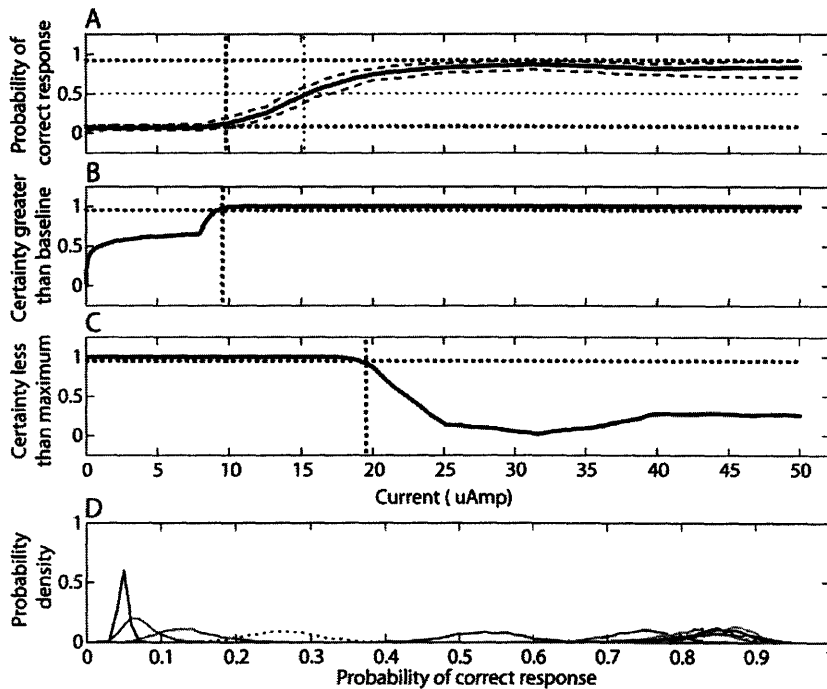
Figure 14: Illustration of inference within Bayesian psychometric function estimation framework. A. Median estimate of the psychometric function model (thick black line) and 95% confidence bounds (dotted lines) sampled at 10 times the rate of the data. Reproduced from Figure 13C. Dotted horizontal line shows the 10, 50 and 90% correct response point and dotted vertical line shows the current level for which the model estimate reaches these values. B. Certainty that the model estimate exceeds the baseline estimate. Certainty computed as the proportion of model sample estimates above the 97.5% bound for baseline. Dotted horizontal line shows the 95% certain line and dotted vertical line shows the current level for which there is 95% certainty that the model estimate exceeds baseline. C. Certainty that the model estimate falls below the maximum model estimate. Certainty computed as proportion of model sample estimates below the 2.5% bound for the maximum model estimate. Dotted lines same as B. D. Probability density of the probability of correct response distributions of model estimates at location of data samples.

In auditory perceptual experiments, threshold, or response threshold (Corso, 1963), is typically defined as the level at which the performance reaches some criterion level, such as 50% correct. We can easily find the current level for which the model reaches 50% correct by finding the level along our estimated psychometric function for which the median (50th percentile) performance estimate equals 50% correct. We call this threshold estimate the *psychometric threshold*. Figure 14A illustrates this procedure visually for the guinea pig drift model. The psychometric threshold occurs at 15.49 and 15.49 re 1μAmp for the guinea pig drift model and no drift model, respectively. Likewise, the threshold occurs at 5.75 and 5.75 μAmp re 1μAmp for the primate drift model and the primate no-drift model, respectively.

The subject's sensitivity to a change in stimulus intensity is often expressed as the width

102

(which we refer to as the *psychometric* width) defined as the range of stimulus intensity values from which the psychometric function changes from α_w to $1 - \alpha_w$. Typically, the range is chosen such that the detection probability ranges from 0.1 to 0.9 (Kuss et al., 2005). However, in the case where asymptotic or best performance does not reach a probability of 1, often described as the subject having a lapse in performance, the absolute selection criterion can be somewhat arbitrary. In the example in Figure 14A of the guinea pig data, maximum performance only reaches 86% correct. For this example, we can find the lower range of the width where performance reaches 10% similarly to the *psychophysical threshold* estimate by finding the stimulus level for which the median model estimate of the probability correct reaches 10%, which is 9.33 μAmp. Because the primate psychometric function exceeds 90% accuracy, we can estimate the *psychometric width* as 3.20 and 3.15 dB μAmp for the drift and no-drift models, respectively.

Another way to compute a threshold is from a purely probabilistic standpoint. We define the *Bayesian threshold* as the level at which we can be reasonably certain (>95%) that the performance at some level is significantly greater than the baseline performance, measured when the stimulus is off (Smith et al., 2004). This definition is often used for inferences in learning literature (Smith et al., 2004; Smith et al., 2005; Smith et al., 2007) and is used for the multi-trial multi-level threshold estimate in Chapter 2. This inference can be made by finding the stimulus level for which at all but the bottom 2.5% of the samples at that level exceed the 97.5% confidence bound for baseline. Figure 14B visually illustrates this inference by showing the certainty of exceeding baseline for the guinea pig drift model. The certainty is defined as the proportion of samples exceeding the upper bound of the baseline model estimate. For the guinea pig data, this Bayesian threshold occurs at 9.33 μAmp for the drift model and at 9.77 μAmp for the no-drift model, and at 4.57 μAmp for the primate data using the drift model and at level 4.62 μAmp for the primate data no-drift model.

As a measure of the width in the probabilistic framework, we would like to suggest that

the range be interpreted as the range of stimulus levels for which performance is significantly different from baseline and best performance. This range would then include the range of levels for which the animal's performance is significantly improving, at a 95% level of significance. The lower bound of the range is the Bayesian threshold, the stimulus levels for which we are 95% certain that performance is better than baseline performance (Figure 14B). The upper bound of the range is the stimulus level for which we can be reasonably certain that the performance is just worse than the best performance. This level of certainty is computed similarly by finding the stimulus level for which the lower 2.5% confidence bound of the model estimate of the maximum performance just falls below 97.5% of the model sample estimates at that stimulus level. Again, we illustrate this procedure for the guinea pig drift model in Figure 14C. The upper bound is estimated as 19.05, 18.62, 7.76, 7.59 μAmp for the guinea pig drift model, no drift model, the primate drift model and the primate no-drift model, respectively. The complete range is the difference between the upper and lower bounds of the range or 9.72, 8.85, 3.19, 2.96 μAmp for the guinea pig drift model, no drift model, the primate drift model and the primate no-drift model, respectively. Of course, the range could be redefined in terms of absolute performance in the same way as we illustrated for computing the psychometric threshold. Additionally, the slope of the psychometric function, here defined as the ratio of performance range of improvement of the psychometric function (in percent) and the range of stimulus levels for which the improvement occurs can be calculated from the above inferences as 5.64, 5.88, 24.96 and 26.12 percent per μAmp for the guinea pig drift, no drift, the primate drift and no drift models, respectively.

Lastly, it is important to emphasize that these, and any, inferences made within this framework are greatly facilitated by the Bayesian estimation procedure. Estimation of the posterior involves estimating many samples from the posterior distribution. Consequently, all inferences are made by computing the proportion of samples within a distribution reaching a certain criterion level. It is therefore useful to show the probability distributions underlying the

inferences. The distributions of the model estimates of the guinea pig drift psychometric model are shown in Figure 14D for the current levels at which the data is collected. These sizes of the distributions are directly reflected in the size of the confidence bounds for the model estimates shown in Figure 14A since the confidence bounds simply encompass 95% of the sampled distributions. It is important to emphasize that we avoid the issue of the hypothesis testing framework in the inference process.

Figure 15: Comparisons of probability of correct detection between two levels shown as the probability that performance at level 1 (x-axis) is greater than performance at level 2. Blue indicates low probabilities while red areas denote high probabilities.
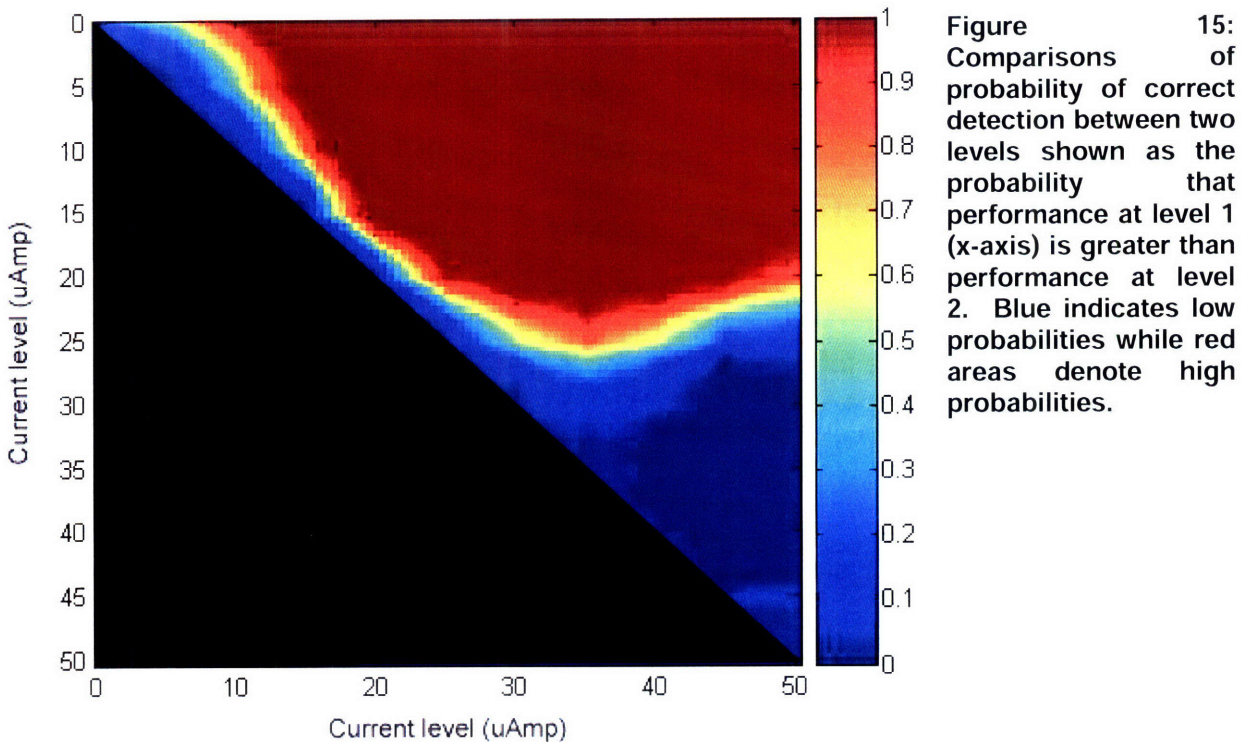
In the discussion of Figure 14, we illustrated comparisons between the performance at one level (the baseline level in Figure 14B and the maximum level in Figure 14C) and all other levels. However, in our framework, we can also compare the subjects' performance between all levels, shown in Figure 15, where we indicate the probability that the performance at one level (x-axis) is greater than the performance at the second level (y-axis). Comparing the performance between the baseline level and other levels in Figure 15 mirrors the findings in Figure 14B. The performance at baseline is smaller than the performance at other levels, but does not become significantly smaller at a level of significance ≥ 0.95 until around 10 µAmp. In fact, Figure 14B is simply one cross-section of Figure 15. Other comparisons indicate that, for

example, performance at 20 µAmp is greater than performance at all current levels below approximately current level 15 µAmp and performance at 40 µAmp is greater than performance at all current levels below approximately 20 µAmp. The shape of the darkest red region indicating a significance level of 0.95 mirrors the shape of the confidence bounds on the estimated psychometric function shown in Figure 14A.

4.1.3. Bayesian model of psychometric function with two or more dependent variables

One benefit of using our stochastic framework for modeling psychometric function is the ability to couple additional parameters to the existing model. For instance, it has been argued in studies of auditory detection as well as other sensory modalities and learning experiments that just as the proportion of correct responses is a proxy for improvements in the detection of some stimulus dimension, so are other proxies such as reaction time (Pfingst et al., 1975a, 1975b; Barnes et al., 2005). In the following section, we illustrate how probability of correct detection and reaction time can be modeled simultaneously in the stochastic framework.

Each of our two datasets (guinea pig and primate) contained not only the proportion of correctly detected trials but also the mean reaction time for each stimulus level, the mean duration of time that the animal spend in making the go/no-go response for all the trials at a given stimulus level. Using the notion that the reaction time is also a proxy our hidden variable of the internal stimulus intensity, x_v, we incorporated reaction time, r_v, by adding another equation to the model as $r_v = b + ax_v + z_v$, where b and a are random variables, and z_v is a zero-mean Gaussian random variable with variance σ_z^{-2}. The priors on b and a are uniform and range between -5 and 5 for a and between 0 and 5 for b. We use gamma priors for σ_z^2 to ensure that the variance of the noise is always positive. The gamma prior is chosen to have a wide range such that the mean is equal to 1 and variance to 0.2. By adding this term, we are able to estimate a model of both the psychometric function (probability of correct detection) and the reaction time simultaneously.
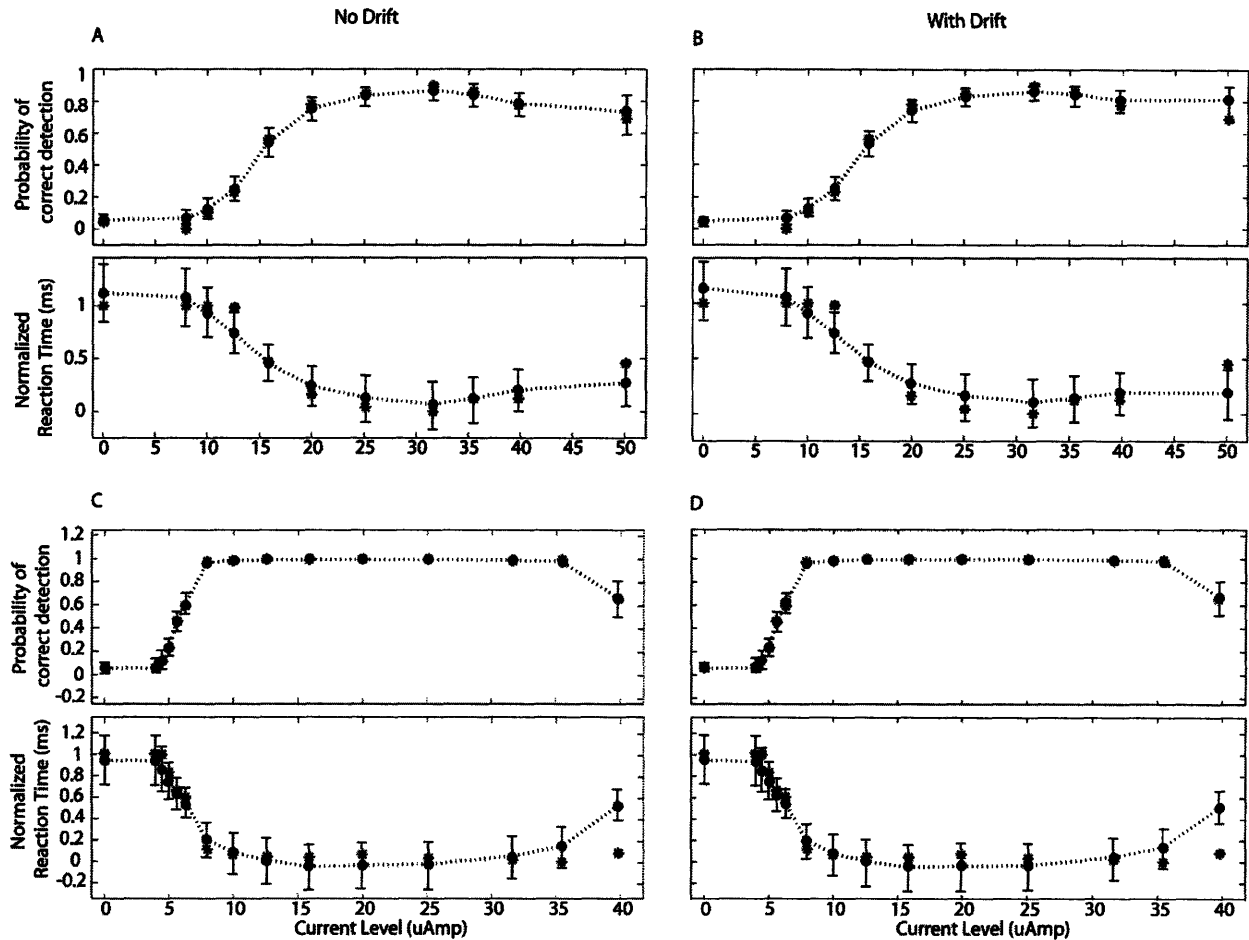
Figure 16: Models with and without drift including reaction time. Left panels show model without the additional drift component (A and C) while right panels include the additional drift component (B and D). Top row models responses for guinea pig (A and B) to trains of brief current pulses, while bottom row models responses of primate (C and D). Reaction times shown are normalized between the minimum and maximum to compare across two species. Gray stars show data, black circles show median of model estimates and error bars show 95% credible intervals.

Figure 16 shows the model estimates for the drift and no drift models for both the guinea pig and primate data incorporating the model of reaction time. These models can be directly compared to models of psychometric functions only shown in Figure 11 and Figure 12. The reaction times are shown and modeled here as normalized reaction times, where the actual mean response data are normalized between 0 and 1. This normalization is done to compare between the two datasets and allow for the same priors. Observing the reaction time data only (gray asterisks), we note that the reaction time mirrors the inverse of the psychometric function. When the probability of correct detection is low, reaction times are high, and when probability of

107

correct detection is asymptotically high, the reaction time is typically asymptotically low. In fact, the lapses in monotonicity seen in the psychometric data (Figure 16A/B highest stimulus level and Figure 16C/D highest stimulus level) are also seen in the reaction times, in this example, although the degree of non-monotonicity differs somewhat between the psychometric and reaction time data.

For the majority of the reaction time data, the model of reaction time, in both the drift and no drift models closely mirrors the data. For instance, the models closely capture the trends in decreased reaction times between 10 and 25 μAmp for the guinea pig data and between 5 and 10 μAmp for the primate data, as well as steadily poor performance below 10 μAmp for the guinea pig data and below 5 μAmp for the primate data. The asymptotic reaction time performance as stimulus levels increase is captured well with some notable exceptions. For instance, although the primate reaction time data at the two highest levels exhibits less non-monotonicity than the percent correct data, both the drift and no drift models are highly influenced by the non-monotonicity at that data point. This divergence from the data directly arises from the assumption that the underlying hidden variable, x_v, internal representation of stimulus intensity, modulates both proxy variables simultaneously and therefore both proxy variable data effect the model estimates of the other.

As was observed for this class of models with probability of correct detection data, the smooth changes in reaction times lead to larger credible intervals for the reaction time data compared to the proportion data. Unlike for the psychometric data, the reaction times change gradually for *both* guinea pig and primate data, and therefore, both sets of data exhibit the larger credible intervals. Similarly to the observations with the psychometric functions, the models tend to smooth out the non-monotonicity in the data, with the drift model incorporating more smoothness than the no drift model. This smoothing can be seen in the guinea pig data for stimulus level of 12 μAmp (Figure 16A/B). The reaction time data at this point is highly non-monotonic, however, the modeled response exhibits monotonicity. This monotonicity is also a

function of the monotonic behavior of the percent correct data, which similarly influences estimates of the underlying hidden level perception parameter that modulates both estimates.

4.1.4. Comparison of inference with models using one and two variables

Just as the psychometric function can be estimated using proportion of correct detections, so too can other variables that serve as proxy variables for the internal representation of the stimulus intensity, such as reaction time, be used to estimate an alternative psychometric function. Therefore, using our dataset, we can estimate the psychometric functions in one of three ways, leading to four estimates of a psychometric function. We can use the traditional proportion of correct detections without reaction times, as shown in Figure 14. Alternatively, we can use the reaction times alone, without the proportion of correct detections to estimate a psychometric function. Third, we can create a model incorporating both proxy variables simultaneously (as shown in Figure 16) to estimate both the psychometric functions based on the proportion of correct detection and the reaction times, and use each for the task of inference. The four possibilities of psychometric functions are shown in Figure 17, with A and C showing psychometric function and certainty plots for models based on one proxy variable (proportion of correct responses and reaction time, respectively) and with B and D showing psychometric functions when both proxy variables are estimated simultaneously. For all plots shown, the sampling frequency is again augmented to ten times the original sampling frequency of the data, as described earlier.
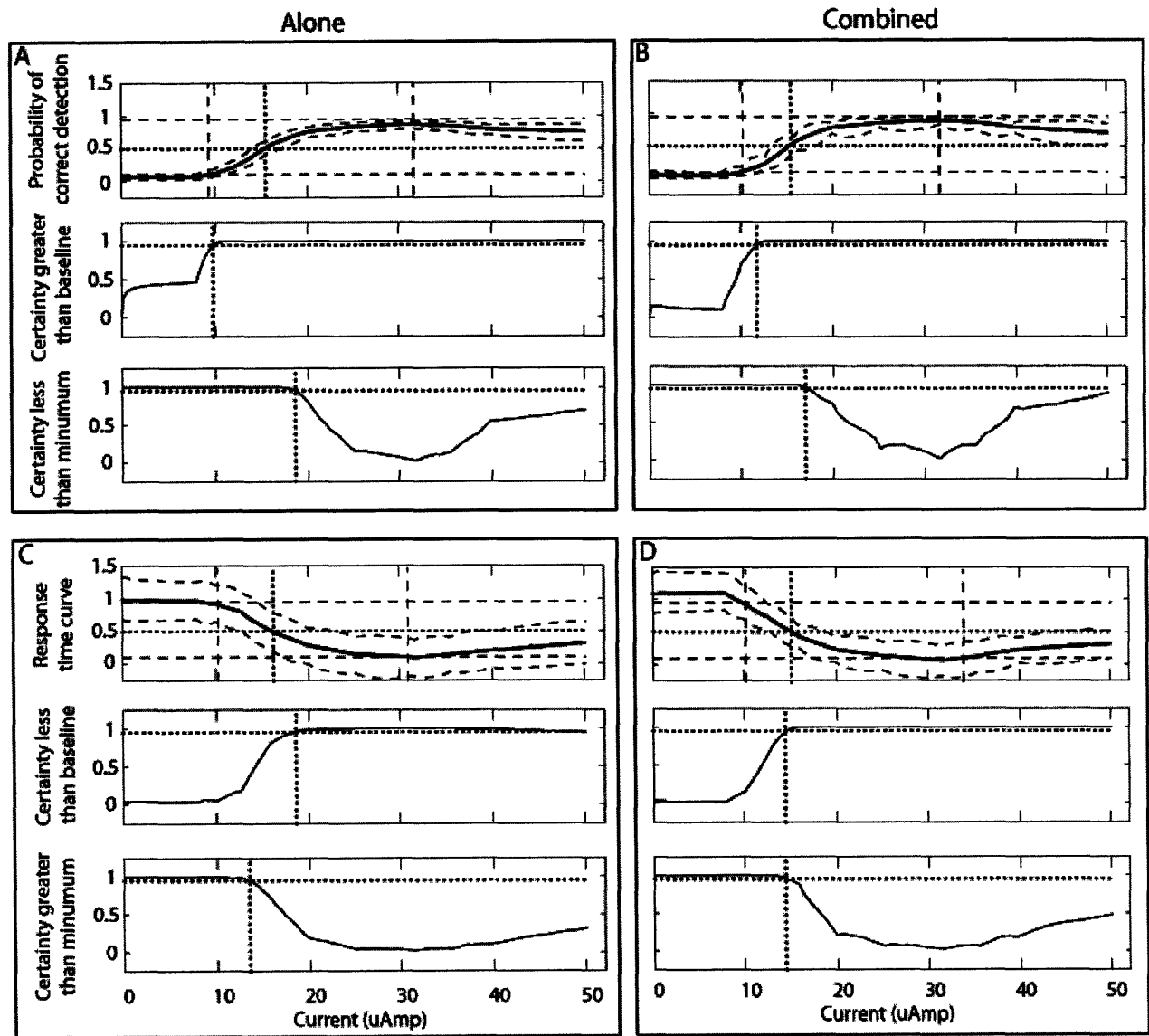
Figure 17: Comparison of inference with 1 and 2 proxy variables. A. Estimation of psychometric function based on the probability of correct detection with 95% credible intervals. The middle plot shows the probability that the performance is greater than baseline at a 95% level of significance. The bottom plot shows the probability that performance is less than the maximum value of the psychometric function at a 95% level of significance. Replicated from Figure 14. B. Estimation of psychometric function based on the probability of correct detection using the number of correct detections and reaction times simultaneously. C. Psychometric function estimates based on reaction time with 95% credible intervals. Middle plot shows the certainty (at a 95% level of significance) that the response at each level is less than the baseline reaction time. Bottom curve shows the certainty (at a 95% level of significance) that the performance is greater than the minimum performance. D. Psychometric function estimates based on reaction time when number of correct detections and reaction times are estimated simultaneously. Guinea pig data shown.

Figure 17A is reproduced from Figure 14 illustrating the model of proportion of correct

responses estimated without reaction times, however Figure 17B shows the analogue using

reaction times alone. As we observed earlier, the reaction times seem to be inversely related to

110

the probability of correct detections, with the animal responding slower for soft stimuli and faster for loud stimuli. We derive the same inferences we outlined for the psychometric function using proportion of correct responses for inferences using the reaction time model. Table 2 compares the inference estimates for models using one proxy variable. Note that because the two proxy variables follow an inverse relationship, the inferences should be compared to reflect this observation. For instance, when comparing the level at which the psychometric function reaches 10% using the proportion of correct responses, the analogous comparison is the level at which the psychometric function reaches 90% using the reaction times. These estimates are similar with the psychophysical threshold (50% detection) estimated as 15.49 µAmp using the proportion of detected responses and 16.22 µAmp using the reaction times. Likewise, 10% correct detection is reached at 9.33 µAmp using proportions of correct responses and 90% of reaction times are found at 10.23 µAmp. However, the differences in the size of the credible intervals for each psychometric function lead to larger differences in estimates of the probabilistic inferences. For instance, the level at which we are 95% certain that performance is greater than baseline is 9.77 µAmp using proportion of correct detections and the level at which we are 95% certain that performance is less than the maximum is 13.49 µAmp using reaction times.

Combining the two proxy variables into one model and estimating both psychometric functions simultaneously could be done when the experimenter places a similar weighting on the information available about the internal representation of the stimulus in both proxy variables. However, unless the two proxy variables move in the same direction for each measurement, this may lead to larger confidence bounds on each psychometric function (compare Figure 17A with Figure 17C and Figure 17B with Figure 17D) and therefore somewhat different estimates of the probabilistic inferences. For example, the level at which we are now 95% certain that performance is greater than baseline using proportion of correct responses has now increased from 9.77 µAmp to 11.15 µAmp and the analogous level for which we are 95%

certain that performance is less than the maximum using reaction times has increased from 13.49 to 14.45 µAmp. Likewise, the levels at which we are 95% certain that performance is greater/less than the maximum/minimum has decreased to 16.98 µAmp and 14.45 µAmp using proportion of correct responses and reaction times, respectively, leading to a reduced probabilistic dynamic range. Therefore, we are now able to make inferences at each level with less certainty, or equivalently, increased performance is needed to estimate with the same degree of certainty when the two proxy variables are combined. However, in certain cases the experimenter may feel that using both proxy variables simultaneously will lead to estimates that more accurately represent the subject's true perception of the stimulus.

	Inference based on probability of correct response (µAmp)			Inference based on reaction time (µAmp)	
	Independent estimation	Combined estimation		Independent estimation	Combined estimation
10% detection	9.33	10.23	90% of reaction time	10.23	10.23
50% detection	15.49	15.49	50% of reaction time	16.22	15.14
90% detection	X	X	10% of reaction time	30.90	33.88
95% certain greater than baseline	9.77	11.15	95% certain performance is less than maximum	13.49	14.45
95% certain less than maximum	18.62	16.98	95% certain performance is greater than minimum	18.62	14.45

Table 2: Summary of inferences shown in Figure 17 for models based on one proxy variable (probability of correct detection or reaction time) and based on both proxy variables estimated simultaneously.

4.2. Random-effects state space model of percentage of activated neurons in two populations

The stochastic models described so far relate the one subject's performance on one task as a function of an independent variable, in this case, stimulus level. These models can be extended to model the performance of multiple subjects in multiple groups, where the subjects

112

in each group are members of some general population. In the following section, we will illustrate and extension of the previously described models to a random effect state space model also implemented in a Bayesian framework using MCMC methods.

4.2.1. Identifying trends in neural responses in the association and no-association subgroups

We developed our RESS model to provide a statistical framework for analyzing proportion data across multiple subjects in multiple groups, such as that used in Barnes et al. (2005). This model could also be used to estimate psychometric functions with multiple subjects in multiple groups. Model estimates for the association group are shown in Figure 18. First, notice that the stochastic smoothing inherent in our model provides relatively smooth estimates of the population response. The modeled population responses are much less variable from learning stage to learning stage because the model assumes that the mean of the hidden variable describing the underlying auditory learning association process is equal to that at the previous level, although allowing for a noise term. This model specification is based on the notion that although the animals only perform the task several times per day, the representation of the learning association remains internally represented from day to day when the animal is not performing the task. This stability of the underlying process supports the notion of stochastic smoothing evidenced in the model estimates. Second, notice that the population responses are not mere averages of the individual subject responses. Rather, the individual responses are drawn from the distribution specified by the population responses. This second observation is based on the notion that all the subjects are members of one population and differences in performance within groups is captured as a random deviation from the population performance.

Our modeling framework, where we estimate all model parameters and their distributions simultaneously, facilitates model inference. The experimenters are interested in inferring trends in the percentage of activated neurons relative to the response during learning stage one. No

113

learning would have occurred at learning stage one since the animal did not have an association between tone and turn at that stage. Therefore, we can compare the performance at the first learning stage with the performance at the other $D-1$ learning stages. Using our MCMC analysis, we can compute the probability that the neural response at learning stage d^* is greater $\Pr(p_{d^*} > p_{d_1})$ or less than $\Pr(p_{d^*} < p_{d_1})$ the baseline response at learning stage d_1. This comparison consists of $D-1$ within-group comparisons. Comparisons for which the probabilities above exceeded 0.90 are considered significant. The computations involve evaluating comparisons from a D-dimentional joint probability density estimated during the MCMC estimation phase. For this reason, the analysis avoids the problem of multiple hypothesis tests. By estimating the distributions and computing the confidence intervals about the estimator (median), we are able to make statistical statements with some pre-determined degree of certainty, unlike in the original analysis of the data performed in Barnes et al. (2005) where no statistical test were used.
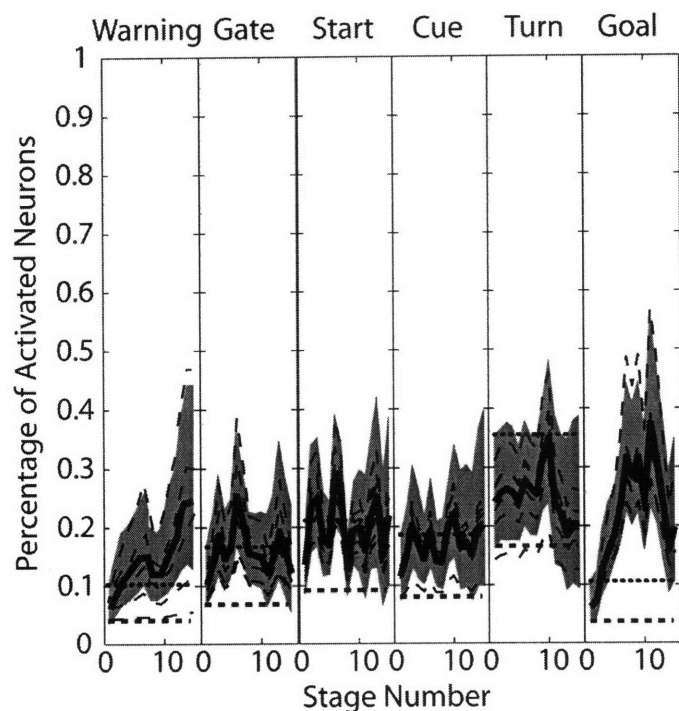
Figure 18: RESS model of association group. Models of the individual animal responses are shown as dashed black lines, the median of the RESS model is shown as a solid thick black line and the gray shaded regions represent the 90% confidence intervals. Dotted line shows the 95% of the baseline response and dashed line shows the 5% of the baseline response. Some individual animal models overlap with median population response. Activated neurons selected by author using methodology outlined in Barnes et al. (2005). Analysis of behavioral responses (not shown) for all the animals in this group indicated that they learned the association between tone and turn .

The dotted line showing the upper 95th percentile of the naïve (first stage) response is

indicated here, useful for making the upward movement inference, and the dashed line showing the lower 5th percentile of the naïve response is also indicated, useful for assessing a downward trend. If a significant pattern exists, then the confidence bounds will be above and below the lines representing 90% certainty for upward and downward movements, respectively. In fact, no evidence for upward or downward trends is seen for the middle four events: gate, start, cue and turn. For these four events, the lack of trend in the population response is most likely attributable to the variation in the individual subject responses both between subjects and as a function of learning stage. In their publication, Barnes et al. (2005) indicated that they observed downward trends for start and turn, contrary to our analysis here. The difference in the findings is most likely related to our computation of confidence intervals about the median. In our models, the bounds are larger for the last several stages of learning since fewer data were available there, illustrating a lesser degree of confidence in the data. However, our analysis does find evidence for an upward trends in the neural responses during warning and goal. This increase is only evident for stages 11 through 15 for warning, and for stages 4 through 14 for goal. There is most likely no significance in increased neural response for the 15th stage of goal because few neurons were represented in the data, since several animals did not complete the last several learning stages in the experiment due to a loss of a headstage. These findings confirm the conclusion in Barnes et al. (2005) that an upward trend is present for warning and gate; however, this analysis specifies with a level of certainty, the stages for which this inference is significant.

We also model the non-association data in the RESS framework to examine whether any within group differences at different learning stages evident as the subjects traversed the T-maze. For these animals, responses to only the first 13 stages of learning were recorded. The RESS models of individual animal responses and the population responses as well as the population 90% credible intervals are shown in Figure 19.
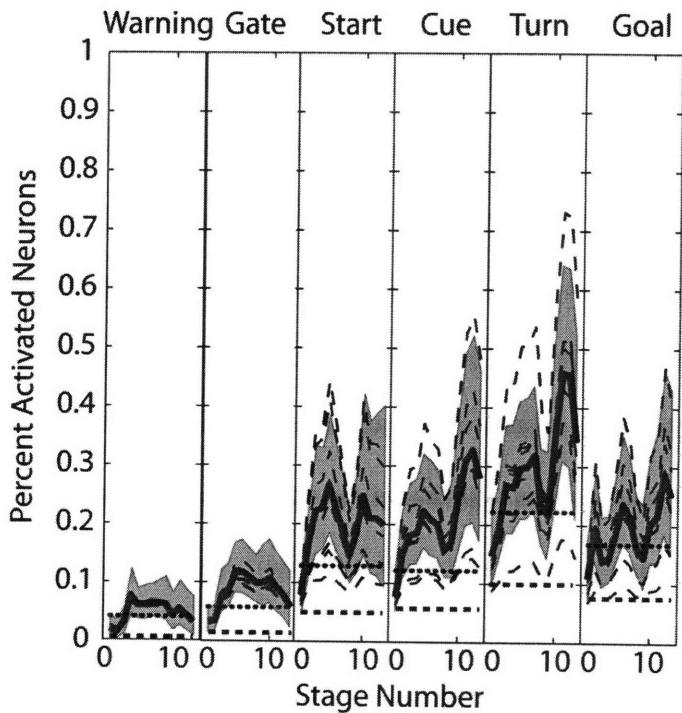
Figure 19: Model of no-association group. Model symbols are same as used in Figure 6. Analysis of behavioral responses (not shown) for all the animals in this group indicated that they did not learn the association between tone and turn .

The events during which within-group effects are evident differ for this no-association group from those seen for the association group. As with the association group, all differences are examined relative to the first (naïve) stage. For both the warning and the goal stages, no increases or decreases in the data could be inferred. The stable neural response during warning stems from the stability in the individual animal responses across all subjects, since the individual model responses perfectly overlap with the population response. However, the stability of the population response for the goal event stem from the variability in both the individual animal responses across animal and across stage. This observation is supported by the wide range in the neural responses for individual subjects.

The population models do not indicate any significant decreases in the responses with learning, however, gate, start, cue and turn responses all exhibit some evidence for increases in proportion of activated neurons for some stages. An increase in the neural response is seen from stage 4 through stage 8 during the gate event, for stage 3 through stage 7 and stage 10 for the start event, for stages 3 through 8 and stages 10 through 13 during the cue event, and for stages 10 through 12 for the turn event. For the gate and start events, there is no indication that

116

the trends continue toward higher learning stages, however, for the cue and turn events, there is some indication of a trend in increased performance for the later stages. Again, there is a great variability in the individual subject responses for the start, cue and turn events, which could underlie the piecewise regions of significance in neural responses between the naïve stage and consequent stages.

4.2.2. Comparing neural responses between the association and no-association subgroups

We also examine whether between group differences in the percent of activated neurons existed. More specifically, we are interested in whether the learning of an association in the association group could be reflected in differences in the percentage of activated neurons between the two groups. That is, we compute the probability that the percentage of activated neurons in the association group is increased over the no-association group, $\Pr\left(p_{d,e}^{Association} > p_{d,e}^{NoAssociation}\right)$, and the probability that the percentage of activated neurons decreased, $\Pr\left(p_{d,e}^{Association} < p_{d,e}^{NoAssociation}\right)$, for learning stages $d = 1, ..., D$ for all six events $e = 1, ..., E$ in the T-maze. We only consider learning stages $d = 1, ..., 13$ because only the first 13 stages of learning were recorded for the no association group. We consider the percentage of activated neurons significantly different if the probability is ≥ 0.90. To conduct this analysis, we again use the confidence bounds of the RESS model to make the between group inference. Visually, this is obtained by overlaying the confidence bounds for all 6 of the events within the T-maze as a function of the learning stage to see if any of the bounds overlap. If no overlap exists for some learning stages for some events, the neural responses signal a between group difference. The between group analysis is shown in Figure 20. We visualize that the only significant group differences exist for the warning event between the two groups. Specifically, the association group shows a significant increase in the proportion of activated neurons relative to the non-association group for all but 4 stages of learning. Between group differences for stages 4, 5, 8 and 9 are not significant at a 90% confidence level, but may be significant at smaller levels of

117

confidence. Although the population averages of the RESS model for other events between the two groups may differ, the differences are not significant at the 90% level of confidence, as shown by the overlap of the confidence bounds for all the other stages and events between the two groups. Therefore, we can conclude that using the percent of activated neurons as a proxy for the neural activity, the responses during the warning event are the only ones that could underlie the neural correlates of learning an association in this experiment.
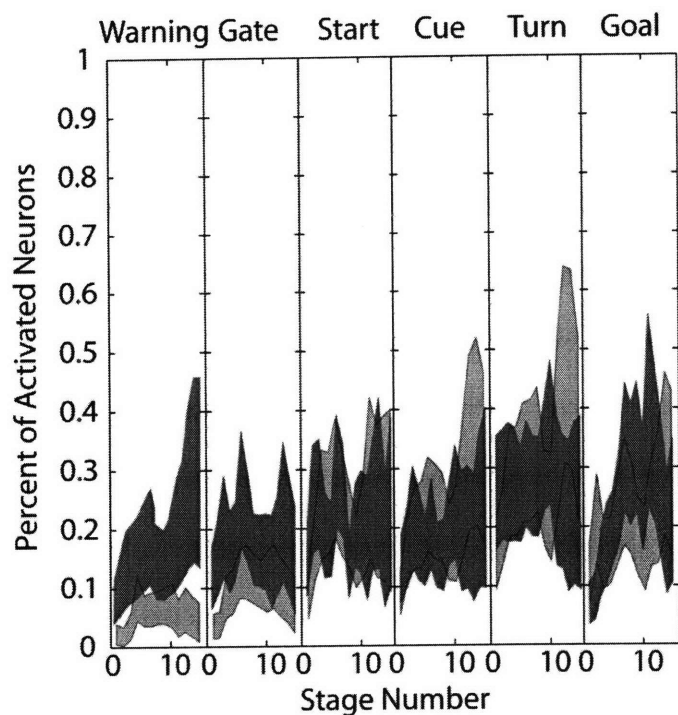
Figure 20: Comparison of between group differences for the association (dark gray) and no-association (light gray) groups at a 90% confidence level. Lack of overlap of confidence regions indicates between group differences. This difference is only seen during the warning period. The credible intervals are the same as those shown in Figure 18 and Figure 19.

4.2.3. Cross-validation analysis of RESS model

We conduct predictive posterior checks testing to assess whether the RESS models fit the underlying data sufficiently well to make inferences. Posterior predictive checks are a generally broad class of goodness-of-fit assessments that evaluate whether new data generated from the model shares similar characteristics with the actual data. A common criterion, which we use here, is the ratio of the mean and variance in the new data (see Methods for further explanation) relative to the original data. Generally, PPC values between .1 and .9 indicate an appropriate data fit (Congdon, 2003) and are generally considered a conservative goodness-of-fit measure (Bayarri and Berger, 2000). We calculate these PPC statistics for our RESS models

118

for both sets of data and results are shown in Figure 21. Most (82/91) of the statistics are centered around a PPC value of .5 and lie between .1 and .9, indicating a sufficient model fit to the data. We therefore conclude that the models fit the data appropriately well and allow for appropriate data inferences.

Figure 21: Predictive Posterior checks for the association dataset (A) and for the non-association dataset (B). The model fits the data well when most of the PPC values lie within .1 to .9, which is the case for this dataset. The different PPC values reflect models fit to individual subjects within each group and for each event.
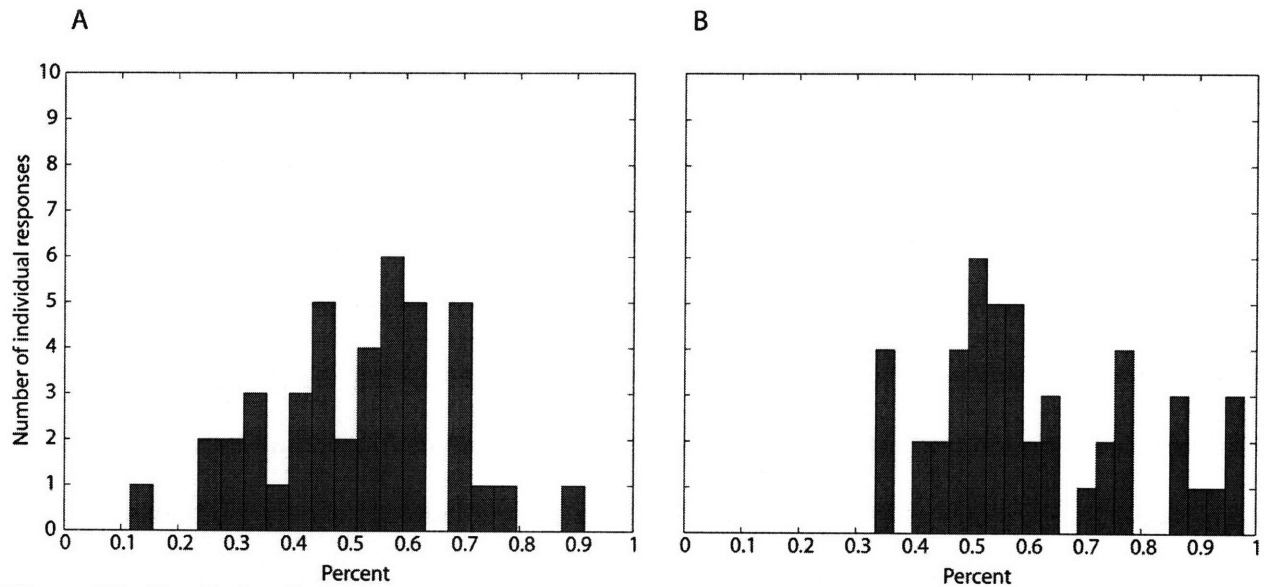
5. DISCUSSION

5.1 Summary

In this work, we illustrate how dynamic stochastic state-space models can be used to represent psychophysical functions, mapping stimulus level to one or multiple metrics of behavioral performance. In estimating psychometric functions, we showed that response threshold and psychometric function width, where both inferences estimate the stimulus corresponding to a specific performance level or range, could be estimated. We also showed that probabilistic inferences of threshold as the stimulus level for which the response at that level is significantly different from baseline, and the width of the range of levels for which the response is different from baseline and maximum performance, can also be estimated in our framework. Although our framework estimates the psychometric function at discrete points, the sampling frequency can be increased allowing the experimenter control over the precision of the estimates. In addition, we illustrated that another measure of behavioral performance, such as reaction time, can be included in the model, such that the joint distribution can be estimated simultaneously for both sets of parameters, and inferences can be made using either or both measurements of behavioral performance. Our state-space model of psychometric functions is extended to a random effect state-space framework when data contains responses from multiple subjects in multiple groups. This RESS model allowed us to quantify population performance of both groups, to characterize the between-subject differences along with assessing between-group differences in the two populations. We used this model to characterize the proportion of activated neurons for animals performing an auditory learning task. Using the RESS model, we inferred that there were significant between group differences in the neural responses during the beginning of the experiment, and that animals in each group showed changes in neural responses throughout learning specific to that group. Furthermore, we compared our inferences to those using a non-model based method and found that our methodology provided stricter assessments of within-group differences.

120

5.2   Advantages of state-space models

The state-space methodology of estimating psychometric functions enjoys several benefits. First, because it facilitates estimating the distributions of the subject's performance for the range of stimuli of interest, including computation of confidence bounds, this methodology can provide insight into inter-subject differences for subjects performing the same task. This aim can be accomplished by estimating psychometric functions for all subjects within our framework. Second, estimating the joint-distribution of the state parameters also allows us compare subject's performance at different stimulus levels by calculating the level of certainty with which performance for one level is greater than the performance at another level. This advantage has prompted us to suggest alternative definitions of threshold based on a reasonable level of certainty that performance at a threshold level is greater than baseline performance in the absence of stimulation to compliment current definitions of threshold that estimate the threshold level as a level for which performance reaches a criterion level. In our framework, the level of certainty can be controlled by the experimenter and altered depending on the specifications of the task. Of course, traditional metrics of threshold based on a specific criterion level of performance (i.e. 50% correct) can also be estimated within our framework.

Third, our framework may be used to provide insight into neural changes that correlate or underlie behavioral performance (Jog et al., 1999; Wirth et al., 2003; Smith et al., 2005) using the same criteria of certainty and estimation methods to model behavioral and neural performance. To evaluate the stimulus for which behavioral performance differs from baseline relative to neural response, the behavioral psychometric function can be estimated using our state-space techniques, while the neural response could be estimated using the state-space techniques outlined in Chapter 2. As an alternative, a previous study estimated the change in the neural activity in a similar framework to the one currently proposed for modeling psychometric functions based on the firing rate (Wirth et al., 2003). Comparisons would be facilitated by both neural and behavioral experiments performed in the same subjects, with the

121

same stimulation levels and using the same criteria (for example 95% certainty) to estimate the threshold levels. In our scenario, the comparison in thresholds between those estimated in Chapter 2 and those estimated here with the guinea pig psychometric function are complicated by the use of different stimuli and different animals. In Lim and Anderson's (2007) data analyzed in Chapter 2, a single current pulse was used to measure responses, while Pfingst and colleagues in this experiment used a series of current pulses to measure behavioral performance. Animals were not able to perform the behavioral task using a single current pulse. However, had the experiments been performed in the same animals and with the same stimulus, the two techniques presented in Chapter 2 and this chapter could have been used to analyze and compare the data concurrently.

The neural responses discussed in this chapter have been analyzed previously using a non-model based method by Graybiel and colleagues (Barnes et al., 2005). The proportions of active neurons were simply averaged across all animals and all neurons for each stage. There are several shortcomings to this method. First, the method does not use any systematic metrics of trends and evaluation of the existence of either an increasing or a decreasing trend is purely qualitative. Second, the methodology did not compensate for uneven numbers of neurons present at each stage. For both the association and no-association group, approximately half of the animals did not complete all stages of the experiment because of a lost or loose headstage and therefore the number of neurons used to calculate the neural performance for the last several stages is typically smaller than those used to calculate neural performance for early stages. However, the non-model method does not take different numbers of neurons at these later stages into account when visualizing trends in neural activity. Third, no statistical tests were used to substantiate the visualized trends and no significance testing was conducted to evaluate the certainty with which a trend exists.

Our RESS framework addresses these shortcomings of the non-model method used by Graybiel and colleagues. Our methodology is model-based and provides a quantitative

systematic assessment of data trends. Specifically, stage-by-stage assessments of the population neural activity are calculated to assess whether the neural activity is increased over baseline. We estimate the joint-posterior probability density of all model parameters and this enables us to construct confidence bounds to quantify our assessments of the certainty of our estimates. Second, our methodology allows for systematic comparisons of trends in neural activity within groups. Because our model compensates for the number of neurons available for the neural response estimates at each stage, the increased confidence bounds reflect stages where fewer neurons were sampled. Consequently, we were only able to detect significant trends in two of the four events identified by the Graybiel lab using their non-model method for the association group (the non-association data was not analyzed by Graybiel and colleagues). Third, by computing the probability that the neural response in one group is greater or less than the neural response in the other group, we could make comparisons of neural activity between the two groups. In doing so, we found significant differences in neural responses only during one event in the T-maze. Our analysis indicated that the proportion of activated neurons is significantly higher for the association task than the non-association task during the warning signal prior to the onset of the experiment during most stages of learning. The neural responses during the other events were not significantly different for the two groups during any stages. Fourth, our between and within-group comparisons were evaluated by estimating from the joint probability density thus avoiding the hypothesis testing.

5.3. Comparisons with current methods of psychometric function estimation

There has been much interest in fitting psychometric functions using constrained continuous models with several parameters. Initial attempts used generalized linear models (Agresti, 1990) to compute ML parameter estimates using the Gumbel/Weibul or Probit functional forms to capture the psychometric function shapes. Since these models map the stimulus level to the [0,1] space, studies suggested that the span of the functional forms would need to be modified to capture a decreased range of behavioral performance. However,

augmentation of the model span, by modeling performance at low levels with a guessing rate and the performance at high stimulus levels as perfect subject to a lapsing rate, made the problems nonlinear and prohibited use of GLM methods (Tretwein and Strasburger, 1999). The field therefore transitioned to ML estimation of the psychometric function defined by two-parameters using a grid-search of the likelihood function (threshold and slope – Watson, 1979; threshold and guessing rate – Green, 1993). However, when fewer parameters than the number necessary to constrain the psychometric function were estimated, the parameters not estimated were set to some value (e.g. zero for lapse rate, chance performance for guessing and something based on prior experiments for slope), introducing bias into all parameters of interest (O'Regan & Humber, 1989). To address this problem, methods were developed to estimate all parameters used to define the psychometric function using nonlinear least-squares estimates with iteratively reweighted nonlinear least square fitting (e.g. Brown & Dennis, 1972). However, the convergence with more parameters was found to deteriorate, and newer techniques have been developed to estimate the four parameters to address this problem in a Bayesian, ML framework (Treutwein and Strasburger, 1999; Kass et al., 2005) and using an empirical framework (Wichmann & Hill, 2001a)

As described, modeling of psychometric functions has focused on estimating continuous valued pre-determined functional forms to fit the behavioral data. When psychometric functional forms are pre-specified (e.g. Treutwein and Strasburger, 1999; Wichmann and Hill, 2001a, 2001b; Kuss et al, 2005), the functional forms are defined in terms of the parameters experimenters seek to infer, such as threshold, width and/or slope and lapse rate. These functional forms lead to parameter estimates that co-vary with each other in the estimation space, meaning that the all parameter estimates will not be independent. This can be undesirable. In our framework, our psychometric functions are not defined in terms of the inference parameters.

Another advantage of our dynamic framework is the lack of assumption about the shape

of the parametric psychometric function. The psychometric functions in our modeling framework are completely data driven and can take on any shape, unlike the sigmoidal functions typically selected by psychophysicists in the current analysis techniques. A consequence of selecting a pre-determined form of the psychometric function can manifest itself in the need to augment the function given different experimental outcomes. For instance, it is generally assumed that the psychometric function maps the stimulus intensity to the [0,1] interval. However, as seen in the data example of the guinea pig discrimination performance, maximum performance levels barely exceed 0.85. No special considerations are needed to fit this dataset in our framework, however, the continuous sigmoidal framework needs to be augmented by adding a lapse term to the data (Wichmann and Hill, 2001a, 2001b). A lapse is defined as a random choice, independent of the stimulus and tends to decrease performance. For instance, in an auditory experiment, a subject may sneeze and miss the presentation of the stimulus, being forced to guess the correct response. To account for lapses, recent techniques suggest adding a lapse probability, π_l, to the model, such that the maximum of the sigmoidal psychometric function changes from 1 to $1 - \pi_l$. If this sigmoidal function estimation is implemented in a Bayesian framework, the choice of priors for the lapse probability may affect the estimation of the psychometric function and inferences. Omission of the lapse parameter may alter threshold inference (Wichmann and Hill, 2001a, 2001b) by introducing large bias and standard deviation for the estimates (Swanson and Birch, 1992). Kuss et al. (2005) discusses that the lapse parameter and psychometric width inference are highly correlated. In the non-stochastic framework, simulation estimation has been developed to reduce, but not eliminate, the bias (Treutwein and Strasburger, 1999).

The current framework also differs from current estimation techniques in that our framework does not force the psychometric function to be non-decreasing. For producing model estimates where the true estimated psychometric function is non-decreasing, we suggest two potential solutions. The first, as illustrated previously, is to introduce a drift term, a, into the

state space equation, as in (3.3) rather than using a simple random walk model as in Eq. (3.2). The added drift term, representing the experimenter's prior notion that the perceived level should increase with larger stimulation level, increases the degree of monotonicity of the estimated psychometric functions as illustrated in Figure 12. However, the estimated psychometric function could remain slightly non-monotonic even including the drift term. A second methodology is to monotonize the data of the observed psychometric function (Ayer et al., 1955) and to use our methodology to estimate the psychometric function using the monotonized data. Briefly, the monotonization technique replaces the ratio of two or more consecutive non-monotonic measurements by a single ratio. For example, if two consecutive values $\hat{p}_i = \frac{n_i}{m_i}$ and $\hat{p}_{i+1} = \frac{n_{i+1}}{m_{i+1}}$ are non-monotonic such that $\hat{p}_i > \hat{p}_{i+1}$, then replace the two values by a single ratio equal to $\hat{p}_i = \hat{p}_{i+1} = \frac{n_i + n_{i+1}}{m_i + m_{i+1}}$. This monotonization technique has been used with the Spearman-Karber method for analyzing psychometric function that also sometimes leads to non-monotonic estimates (Miller and Ulrich, 2001).

5.4 Bayesian formulation approach

The models presented in this chapter could have been estimated by maximum likelihood using EM algorithm (Smith et al., 2004, 2005). However, because of the empirical nature of those techniques, one would need to rewrite the algorithms whenever model specifications are changed. In contrast, application of Monte Carlo Markov Chain methods allows us to conduct the model fitting in winBUGS with the software formulating efficient sampling methods to estimate the exact posterior densities of the parameters, rather than rewriting our own algorithms. Additionally, asymptotic Gaussian considerations were employed to solve the EM algorithms for this class of models (Czanner et al., 2008; Smith et al. 2003; Chapter 2), while the sampling distributions here are exact. On the other hand, Bayesian techniques necessitate specifying prior parameter distributions. However, Bayesian methods can handle the presence of prior information, with inclusion of both subjective and objective priors, although the

robustness of the posterior estimates with respect to the priors should be ascertained (Ghosh et al., 2006). Smith et al. (2007) shows that for this class of problems, model inferences using EM and Bayesian techniques virtually do not differ.

Additionally, Bayesian sampling techniques facilitate inference estimation over EM methods. Using EM methods, we first solved for the parameter estimates and computed the covariance matrix to estimate the distribution of the parameters. Next, we used Monte Carlo techniques to estimate confidence bounds using our estimates of the parameters and the covariance matrix (see Chapter 2; Smith et al. 2003; Czanner et al., 2008). To compare between two groups, in the case of the population learning, we would have needed to draw samples from the approximately Gaussian distributions of our parameter estimates (using the parameter estimates and covariance matrix estimates), and calculate the probability that samples from one distribution exceeded samples from another distribution. In the Bayesian sampling case, no further Monte Carlo simulations are needed, since the MCMC schemes sample from the underlying distributions. By comparing the sampled parameter distributions from the two groups, we can easily assess whether significant between-group trends exist. Further, because of the probabilistic nature of the parameter and credible interval interpretations, Bayesian methods often allow for easier interpretability (Ghosh et al., 2006).

ACKNOWLEDGEMENTS

6. REFERENCES

Agresti A (1990) *Categorical data analysis*. New York: Wiley.

Ayer M, Brunk HD, Ewing GM, Reid WT, Silverman E (1955) An empirical distribution function for sampling with incomplete information. *Ann Math Stat* 26: 641-647.

Barnes TD, Kubota Y, Hu D, Jin DZZ, Graybiel AM (2005) Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature* 437: 1158-1161.

Blackwell HR (1952) Studies of psychophysical methods for measuring visual thresholds. *J Opt Soc Am*, 42, 606-616.

Brown KM and Dennis JEJ (1972) Derivative free analogues of the Levenberg-Marquardt and Gauss algorithms for nonlinear least squares approximation. *Numerische Mathematik* 18: 289-297.

Bayarri MJ, Berger JO (2000) P values for composite null models. *J Am Stat Assoc* 95:1127-1142.

Cappé O, Moulines E, Rydén T (2005) *Inference in hidden Markov models.* New York: Springer.

Congdon P (2003) *Applied Bayesian Modelling.* John Wiley and Sons Ltd., Chichester, United Kingdom.

Corso JF (1963) A theoretico-historical review of the threshold concept. *Psychol Bull* 60, 356-370.

Czanner G, Eden UT, Wirth S, Yanike M, Suzuki W, Brown EN (2008) Analysis of within-and between-trial neural spiking dynamics. *J Neurophysiol* In Press.

Dougherty RF, Cynader MS, Bjornson BH, Edgell D, Giaschi DE (1998) Dichotic pitch: a new stimulus distinguishes normal and dyslexic auditory function. *Neuroreport* 9:3001-3005.

Efron B, Tibshirani RJ (1993) *An introduction to the bootstrap.* New York: Chapman & Hall.

Fahrmeir L, Tutz D (2001) *Multivariate statistical modeling based on generalized linear models.* New York: Springer-Verlag.

Friesen LM, Shannon RV, Beskent D, Want X (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. J Acoust Soc Am: 110: 1150- 1163.

Fu QJ, Shannon RV (1998) Effects of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners. J Acoust Soc Am 104: 2570-2577

Gamerman D (1997) *Markov chain monte carlo: stochastic simulation for Bayesian inference.* London: Chapman & Hall.

Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Machine Intell*, PAMI-6:721-741.

Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc*, 85: 398-409.

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences [with discussion]. *Stat Sci* 7: 457-511.

Gelman A, Carlin JB, Stern HS, Rubin DB (1995) *Bayesian data analysis*. New York: Chapman & Hall/CRC.

Ghosh JK, Delampady M, Samanta T (2006) *An introduction to Bayesian analysis: theory and methods*. New York: Springer.

Gilks WR, Richardson S, Spiegelhalter DJ (1996) *Monte Carlo Markov chain in practice*. New York: Chapman and Hall/CRC.

Green, DM (1993) A maximum-likelihood method for estimating thresholds in a yes-no task. *J Acoust Soc Am* 87: 2096-2105.

Hughes ML, Vander Werff KR, Brown CJ, Abbas PJ, Kelsay DMR, Teagle HFB, Lowder MW (2001) A Longitudinal Study of Electrode Impedance, the Electrically Evoked Compound Action Potential, and Behavioral Measures in Nucleus 24 Cochlear Implant Users. *Ear Hear* 22(6):471-486.

Jog MS, Kubota Y, Connoly CI, Hillegaart V, Graybiel AM (1999) Building neural representations of habits. *Science* 286: 1745-1749.

Kass RE, Carlin BP, Gelman A, Neal RM (1998) Markov Chain Monte Carlo in practice: a roundtable discussion. *Am Stat* 52: 93-100.

Kitagawa G, and Gersh W (1996) *Smoothness priors analysis of time series*. New York: Springer.

Koch KR (2007). *Introduction to Bayesian statistics*. New York: Springer.

Kuss M, Jakel F, Wichmann FA (2005) Bayesian inference for psychometric functions. *J Vis* 5: 478-492.

Laird NM, Ware JH (1982) Random-effects models for longitudinal data. *Biometrics* 38: 963-974.

Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS-a Bayesian modeling framework: concepts, structure and extensibility. *Stat Comput* 10:325-337.

Miller AL, Morris DJ, Pfingst BE (2000) Effects of time after deafening and implantation on guinea pig electrical detection thresholds. *Hear Res* 144: 175-186.

Miller AL, Smith DW, Pfingst BE (1999). Across-species comparisons of psychophysical detection thresholds for electrical stimulation of the cochlea: II. Strength-duration functions for single, biphasic pulses. *Hear Res* 135: 47-55.

Miller J, Ulrich R (2001) On the analysis of psychometric functions: The Spearman-Karber method. *Percept Psychophys* 63: 1399-1420.

Murphy K, Mahdaviani M (2005) MATBUGS software. http://www.cs.ubc.ca/~murphyk/-Software/MATBUGS/matbugs.html

O'Regan JK, Humbert R (1989) Estimating psychometric functions in forced choice situations: Significant biases found in threshold and slope estimations when small samples are used. *Percept Psychophys* 45: 434-442.

Pfingst BE, Colesa DJ, Bledsoe Jr. SC, Benson JM (2007) Psychophysical and neurophysiological studies of chronic implants in the inferior colliculus: Comparison to cochlear implants. *Assoc Res Otolaryngol Abst* 30, 82.

Pfingst B., Glass I, Spelman FA, Sutton D (1985) Psychophysical studies of cochlear implants in monkeys: Clinical implications. In: Schindler RA, Merzenich, MM, (Eds.), Cochlear Implants. Raven Press, New York. pp. 305-321.

Pfingst BE, Hienz R, Miller J (1975a) Reaction-time procedure for measurement of hearing. I. Suprathreshold functions. *J Acoust Soc Am* 57: 421-430.

Pfingst BE, Hienz R, Miller JM (1975b) Reaction-time procedure for measurement of hearing. II. Threshold functions. *J Acoust Soc Am* 57, 431-436.

Rance G, Dowell RC, Rickards FW, Beer DE, Clark GM (1998) Steady-state evoked potential and behavioral hearing thresholds in a group of children with absent click-evoked auditory brain stem responses. *Ear Hear* 19: 48-61.

Smith AC, Brown EN (2003) Estimating a state-space model from point process observations. *Neural Comp* 15: 965-991.

Smith AC, Frank LM, Wirth S, Yanike M, Hu D, Kubota Y, Graybiel AM, Suzuki W, Brown EN (2004) Dynamic analysis of learning in behavioral experiments, *J Neurosci* 15: 965-91.

Smith AC, Stefani MR, Moghaddam B, Brown EN (2005) Analysis and design of behavioral experiments to characterize population learning. *J Neuophysiol* 93: 1776-1792.

Smith AC, Wirth S, Suzuki W, Brown EN (2007) Bayesian analysis of interleaved learning and response bias in behavioral experiments. *J Neurophys* 97: 2516-2524.

Spiegelhalter DJ, Thomas A, Best N, Lunn D (1996-2004). WinBUGS v. 1.4.1 Imperial College and Medical Research Council (MRC), United Kingdom. http://www.mrc-bsu.cam.ac.uk/bugs/

Swanson WH, Birch EE (1992) Extracting thresholds from noisy psychophysical data. *Percept Psychophys* 51: 409-422.

Stiratelli R, Laird N, Ware JH (1984) Random-effects models for serial observations with binary response. *Biometrics* 40: 961-971.

Treutwein B, Strasburger H (1999) Fitting the psychometric function. *Percept Psychophys* 61: 87-106.

Watson AB (1979) Probability summation over time. *Vision Res* 19: 515-522.

Werner LA, Marean GC, Halpin CF, Spetner NB, Gillenwater JM (1992) Infant auditory temporal acuity: gap detection. *Child Dev* 63: 260-272.

Wichmann FA, Hill NJ (2001a). The psychometric function. I. Fitting, sampling and goodness-of-fit. *Percept Psychophys* 63: 1293-1313.

Wichmann FA, Hill NJ (2001b). The psychometric function. II. Bootstrap-based confidence intervals and sampling. *Percept Psychophys* 63: 1314-1329.

Youngentob SL, Margolis FL (1999) OMP gene deletion causes an elevation in behavioral threshold sensitivity. *Neuroreport* 10: 15-19.

Chapter 4

Improved discrimination performance of auditory responses to natural stimuli due to use of point process techniques

1. ABSRACT

Neural spiking activity is a dynamic stochastic process affected by multiple factors, including the extrinsic stimulus dynamics, the neuron's own spiking history and ensemble activity from nearby neurons. Because action potentials are considered all-or-nothing events, the response can be modeled as a point process, a binary series of impulses. However, many current models of auditory spike trains do not treat the spikes as a point process. Instead, by averaging across multiple stimulus presentations, or smoothing and binning, they transform the spikes from a point process into continuous-valued data. Here, we develop a point process generalized linear model framework to analyze responses of sensory neurons to natural stimuli. The framework can accommodate an arbitrary number of covariates, assesses their relative importance and utilizes goodness-of-fit metrics appropriate for point process data. We illustrate our approach with models of single-unit activity recorded from the auditory nerve and cochlear nucleus of cat in response to a spoken sentence (Delgutte et al., 1998) and from field L of a male zebra finch (the avian homolog to the primary auditory cortex) in response to 20 conspecific bird songs (Narayan et al., 2006). We represent the natural stimulus dynamics using a mode decomposition approach suitable for representation for non-stationary signals and show that it provides a parsimonious decomposition with a high degree of goodness-of-fit for all three recording sites. Assessing the history dependence in the auditory nerve and the cochlear nucleus, we find that the neurons exhibits refractory behavior until 14 ms after the previous spike, with greater suppression on average seen in the cochlear nucleus than in the auditory nerve. We also perform a decoding analysis to quantify the degree to which a single neuron in field L can discriminate between conspecific songs. We find that our point process likelihood-

based framework allows for levels of discrimination that are similar to those observed behaviorally in the same species, without smoothing or filtering the original spike data. These findings have important implications for developing theoretically-sound and practical descriptions of neural responses and for understanding information transmission within the auditory system.

2. INTRODUCTION

Investigators have attempted to understand the specificity of auditory neural responses from virtually the time that single unit recordings became feasible. Understanding what factors cause neurons in the auditory system to spike has been an extremely challenging problem, and one that remains largely unsolved. It is known that multiple factors simultaneously influence the spiking of single neurons; however the identity of these factors and their relative importance is largely unknown. Early pioneers in single unit recordings studied responses in the peripheral auditory nerve, the first processing station in the auditory system. By first presenting tone pips at various frequencies and intensities (Kiang, 1965; Sachs and Kiang, 1968; Kiang and Moxon, 1973; Sachs and Abbas, 1974; Liberman, 1978) and later speech and other more complex sinusoidally amplitude modulated stimuli (Young and Sachs, 1979; Sachs and Young, 1979; Delgutte, 1980; Delgutte and Kiang, 1984; Sinex and Geisler, 1981; Javel, 1980), investigators began to theorize that the auditory nerve responds to sound intensity within a band-limited range of frequencies. Later studies explored more central nuclei in the auditory system, such as those in the superior olive. Investigators have concluded that the medial superior olive, which receives input indirectly from each auditory nerve, is largely responsible for our ability to interpret minute temporal differences between the two ears (Masterton et al., 1967; Goldberg and Brown, 1969; Yin and Chan, 1990), and that the lateral superior olive is largely responsible for interpreting small changes in stimulus intensity between the two ears (Sanes and Rubel, 1988; Sanes, 1990; Carr and Konishi, 1990). Functions of other monaural and binaural nuclei within the central auditory system have been postulated. For instance, a modulation frequency map, (Schreiner and Langner, 1988; Langner and Schreiner, 1988) and coding of musical dissonance (McKinney et al., 2001) in the inferior colliculus, pitch coding neurons in the primary auditory cortex (Kadia and Wang, 2003; Bendor and Wang, 2005) and novelty detection in auditory forebrain (Gill and Theunissen, 2007); however, deterministic and all-encompassing theories about the role of each nucleus in the auditory system do not yet exist. In general, the

134

functionality of more central neurons is more uncertain than those more peripheral.

Current spiking history is also strongly influenced by neuron's own past spiking. Previous spiking history reflects the biophysical properties of the neuron such as refractoriness, rebound excitation or inhibition (Truccolo et al. 2005) and rhythmic or network dynamics (Czanner et al., 2008). Studies have attempted to characterize these biophysical properties in the auditory nerve (Gaumond et al., 1982; Johnson and Swami, 1983; Mark and Miller, 1992) and lateral superior olive (Johnson et al., 1986; Johnson et al., 1986; Johnson, 1996); however, these studies have not presented the statistical significance of the effect of past spiking on current spiking propensity, have restricted their analysis to refractoriness behavior without examining other biophysical neural properties for longer interspike durations, such as rebound excitation or inhibition, and have not characterized the effect of past spiking for other auditory nuclei.

Current attempts to understand the relationship between spiking activity and factors affecting the spiking activity have generally focused on linear methods or nonlinear non-regression phenomenological models. Early attempts to estimate only the stimulus parameters that initiated an action potential focused on the method of reverse correlation (de Boer and Kuyper, 1968), by averaging the portion of the stimulus before each spike to obtain the neural spike triggered average. These and consequent methods apply linear filtering theory to derive the optimal linear filter that predicts the post-stimulus time histogram of the response from the stimulus, first suggested by Aertsen and Johannesma (1981). This interpretation of the STRF has gained great popularity and has been applied to several auditory centers in various species (e.g. Theunissen et al., 2000; Narayan et al., 2006; Depireux et al., 2001; Escabi et al., 2003). Phenomenological models are typically nonlinear and nonregression-based and study the possible neural mechanisms underlying cellular changes that elicit action potentials in various auditory centers (Zacksenhouse et al., 1998; Wang et al., 1992; Carney, 1993; Zhang et al., 2001; Kalluri and Delgutte, 2003a; Kalluri and Delgutte 2003b).

135

Although the current methodologies have provided important characterizations of neural system response properties, they possess several shortcomings. First, the neural spike trains are not always considered a sequence of discrete events that form a point process time series (Brillinger 1988). Linear or continuous time methods are used to analyze the spike trains, and as a consequence, the spike trains are either smoothed, binned, averaged across trials or all of the above. These preprocessing steps can strongly influence the structure of the models and affect the inferences of the analysis. Second, because current methods do not always isolate the previous spiking history, they confound the effects of past neural spiking and do not allow the model to assess the relative importance of previous biophysical events. In addition, goodness-of-fit metrics and decoding analysis to evaluate the model performance in capturing the neural response are rarely done (Brown et al., 2002; Truccolo et al., 2005).

We offer an alternative representation to characterize auditory spiking activity. Our alternative model presents the transformation from stimulus to spikes in a point process likelihood framework, offering a broad class of models to examine various stimulus-dependent and biophysical aspects of the response. This framework has been used to characterize neural responses at the level of individual spike times in the auditory (Dreyer et al., 2007; Czanner et al., 2008), visual (Brown et al., 2002), somatosensory (Haslinger et al., 2007), motor (Brillinger, 1998; Eden et al., 2004; Paninski, 2004; Truccolo et al., 2005) and memory systems (Brown et al., 1998, Okatan et al., 2005; Brown et al., 2002; Eden et al., 2004; Czanner et al., 2008), as well as to differentiate between response properties in normal and diseased patients (Sarma et al., 2008). Unlike current methods that sometimes ignore the stochastic nature of neural responses and average responses across trials, our methodology models the underlying action potential spike trains and treats them as a point process. This dynamic model of spikes captures the multiple factors influencing the spiking activity of neurons and assesses the importance of each factor in determining the relationship between the stimulus and the spike output. In addition, our framework allows for goodness-of-fit and decoding metrics appropriate

for point process models. These metrics assess the degree to which the model serves as a good description of the data, that the model can predict the neural response on a spike-by-spike basis and that the model covariates can be removed without inappropriately deteriorating the model fit.

To illustrate our methods, we model responses to natural stimuli within three nuclei in the auditory system. We model the responses to a sentence in the mammalian auditory nerve and cochlear nucleus (Delgutte et al., 1998), and responses to bird songs in field L, the avian homolog to the primary auditory cortex, of a male zebra finch (Wang et al., 2007). Our models capture the effect of past spiking history and of decompositions of the stimulus. To decompose the stimulus, we employ a recently-proposed methodology, the empirical model decomposition (Huang et al., 1998), appropriate for analyzing non-stationary, non-linear signals. The decomposition extracts the natural time-scales present in the stimulus and is a compact representation of the stimulus. Using these covariates, we first show that our framework provides an excellent representation of the data and predicts the data on a spike-by-spike basis. Second, we illustrate the interpretability of our parameter estimates and how these estimates can be used to gain greater insight into the signal processing of the cell. Third, we conduct decoding analysis on the bird song dataset and illustrate that our method is able to discriminate between neural responses to different songs with precision comparable to that seen behaviorally.

3. METHODS

3.1 General likelihood formulation of point process model

Action potentials are transmitted from cell to cell as very rapid increases in the cell's membrane voltage, occurring on the order of microseconds. Neuroscientists regard action potentials as all or nothing events, where the timing and rate, but not the size of the action potential, matter for information transmission. Therefore, we assume that spikes can be modeled as a point process, a time-series of binary events occurring in continuous time, where the binary events represent a spike (1) or no spike (0) (Brown et al, 2003; Daley and Vere-Jones, 2003; Brown, 2005). On a recording interval of $(0, T]$, the counting process $N(t)$ represents the J spike times that have occurred on the interval $(0, t]$ at times $0 < u_1 < u_2 < u_3 < \cdots < u_J < T$. Consequently, a model of the point process can be completely characterized by the conditional intensity function (CIF) that defines the instantaneous firing rate at every point in time as:

$$\lambda(t|H(t)) = \lim_{\Delta \to 0} \frac{\Pr\left[N(t+\Delta) - N(t) = 1|H(t)\right]}{\Delta} \qquad (4.1)$$

where $H(t)$ represents the spiking history until time t (Brown et al., 2003; Daley and Vere-Jones, 2003; Brown, 2005). This is a generalization of the definition of a Poisson point process to one that is inhomogeneous and history dependent.

We formulate a specific form of the conditional intensity function to reflect common paradigms of auditory neural data collection where data are collected over $1, \ldots, K$ trials in response to a sound stimulus. We discretize our representation of time on the duration $[0, T)$ into B intervals, such that the length of each time interval is $\Delta = TB^{-1}$ and each time step is indexed as $b\Delta$. In this case, the discrete version of the conditional intensity function at time $b\Delta$ is $\lambda_k(b\Delta|H_{k,b})$, where $H_{k,b}$ represents the vector of spiking history at trial k up to time bin b. We select a sufficiently large B so that $\lambda_k(b\Delta|H_{k,b})\Delta$ approaches the probability of seeing one

spike in the time interval Δ. We can also represent the parametric form of the conditional intensity function at time $b\Delta$ as $\lambda_k\left(b\Delta \middle| \Psi, H_{k,b}\right)$ where Ψ represents parameters to be estimated. We represent the spike times as a series of indicator functions $n_{k,b}$ that take on the value of 1 if a spike occurs in the interval $((b-1)\Delta, b\Delta]$ for trial k, and takes on a value of 0 otherwise. We represent the binned spike times at trial k as $N_k = \{n_{k,1}, \ldots, n_{k,B}\}$ and represent the entire dataset as $N_{1:K} = \{N_1, \ldots, N_K\}$.

We use likelihood methods for fitting the conditional intensity function to the data. To use these methods, one must derive the joint probability density for the spike train. If we select a sufficiently small Δ such that the probability of seeing more than one spike in this binwidth approaches 0, the joint probability can be written as the product of Bernoulli independent events (Truccolo, et al., 2005):

$$
\begin{aligned}
\Pr[N_{1:K}|\Psi] = \prod_{k=1}^{K}\prod_{b=1}^{B} &\left[\lambda_k\left(b\Delta|H_{k,b},\Psi\right)\Delta\right]^{n_{k,b}} \\
* &\left[1 - \lambda_k\left(b\Delta|H_{k,b},\Psi\right)\Delta\right]^{1-n_{k,b}} + o(\Delta)
\end{aligned}
\tag{4.2}
$$

where $o(\Delta)$ represents the probability of seeing two or more spikes on the interval $((b-1)\Delta, b\Delta]$. If we select Δ sufficiently small (for example, the refractory period of a neuron) such that the occurrence of two or more spikes is highly unlikely, then this probability $o(\Delta)$ tends to zero:

$$
\lim_{\Delta \to 0} \frac{o(\Delta)}{\Delta} \to 0.
\tag{4.3}
$$

For very small Δ, $\left[1 - \lambda_k\left(b\Delta|H_{k,b},\Psi\right)\Delta\right] = \exp\left(-\lambda\left(b\Delta|H_{k,b},\Psi\right)\Delta\right)$ and $\log\left[\left(\lambda_k\left(b\Delta|H_{k,b},\Psi\right)\Delta\right)\left(\left(1 - \lambda_k\left(b\Delta|H_{k,b},\Psi\right)\right)\Delta\right)^{-1}\right] \approx \log\left(\lambda_k\left(b\Delta|H_{k,b},\Psi\right)\Delta\right)$. Making these approximations, (4.2) can be reduced to:

$$\text{Pr}[N_{1:K}|\Psi]$$

$$= \prod_{k=1}^{K}\prod_{b=1}^{B}\left(\frac{\lambda_k(b\Delta|H_{k,b},\Psi)\Delta}{1-\lambda_k(b\Delta|H_{k,b},\Psi)\Delta}\right)^{n_{k,b}}\prod_{k=1}^{K}\prod_{b=1}^{B}\exp\left(-\lambda(b\Delta|H_{k,b},\Psi)\Delta\right)+o(\Delta). \qquad (4.4)$$

Making the assumption that the binsize is small enough such that $o(\Delta) \to 0$, we can rewrite (4.4) as:

$$\text{Pr}[N_{1:K}|\Psi] = \exp\left(\sum_{k=1}^{K}\sum_{b=1}^{B}\log[\lambda_k(b\Delta|H_{k,b},\Psi)\Delta]\,n_{k,b}\right.$$

$$(4.5)$$

$$\left.-\sum_{k=1}^{K}\sum_{b=1}^{B}\lambda_k(b\Delta|H_{k,b},\Psi)\Delta\right).$$

If we view (4.5) as a function of Ψ, then this equation represents the likelihood function of the discrete-time point process, $L(\Psi|N_{1:K})$. Discretization of the CIF is not necessary and the joint probability density of the point process spike train can be defined in continuous time (Brown et al., 2003; Daley and Vere-Jones, 2003). For sufficiently small Δ, the likelihood function in (4.5) is described by the Poisson probability mass function (PMF) and, can also be represented by a Bernoulli distribution, since the two are equivalent in this case. If the logarithm is expressed as a function of linear combinations of the covariates as:

$$\log[\lambda_k(b\Delta|H_{k,b},\Psi)\Delta] = \sum_{r=1}^{R}\theta_r g_r(b\Delta), \qquad (4.6)$$

where $g_r(b\Delta)$ are general basis functions and R represents the dimensionality of the estimated parameter θ, then the PMF has the form of a likelihood for a generalized linear model (GLM) under a Poisson probability model and a log link function (Truccolo, 2005; Czanner et al., 2008). The model parameters can be efficiently estimated by maximum likelihood using the iteratively reweighted least squares algorithm (McCullagh and Nelder, 1989).

3.2 Adopted model

We assume that the model for the CIF $\lambda_k(b\Delta|H_{k,b},\Psi)$ describes both the dynamics related to the previous spiking history as well as the portion of the response directly related to

the stimulus dynamics. We define the cell's previous spiking history as $\lambda_k^H\left(b\Delta|H_{k,b},\gamma\right)$ and the response to the stimulus dynamics as $\lambda_k^S(b\Delta|\theta)$. The stimulus dynamics describes the cell's response to different stimulus components controlling for the previous spiking history. Therefore, we express the complete CIF as a product of the stimulus and past spiking history components of the response (Gaumond et al., 1982):

$$\lambda_k\left(b\Delta|H_{k,b},\Psi\right) = \lambda_k^S(b\Delta|\theta)\lambda_k^H\left(b\Delta|H_{k,b},\gamma\right). \tag{4.7}$$

To model the stimulus effect of the instantaneous firing rate, we define the following form for the stimulus-related component of the CIF in (4.7) :

$$\lambda^S(b\Delta|\theta) = \exp\left(\sum_{r=1}^{R} \theta_r g_r(b\Delta)\right), \tag{4.8}$$

where $g_r(b\Delta)$ are general basis functions and R represents the dimensionality of the estimated parameter $\theta = \{\theta_1, \dots, \theta_r, \dots, \theta_R\}$. For this dataset, we assume that the trial variations can be captured in the history component of the response and we therefore assume a trial-independent parameterization of the stimulus parameter θ. The form of the $g_r(b\Delta)$ basis functions can be quite general.

We assume the following form for the component of the CIF in Eq. (4.7) describing the past spiking history:

$$\lambda_k^H\left(b\Delta|H_{k,b},\gamma\right) = \exp\left(\sum_{j=1}^{J} \gamma_j n_{k,b-j}\right), \tag{4.9}$$

where γ_j relates the current spiking activity to spiking activity during the previous J intervals. We capture the history dependence $H_{k,b\Delta}$ at a resolution of Δ, which is typically 1 ms consistent with the refractory period of a neuron. Past spiking history can capture observed phenomena such as self-inhibiting behavior (refractory period), self-exciting behavior and bursting and rhythmic or network effects.

Therefore, combining equations (4.7) (4.8) and (4.9) , we obtain the following complete

form of the CIF for our problem:

$$\lambda_k\left(b\Delta|\theta,\gamma,H_{k,b}\right) = \exp\left(\sum_{r=1}^{R}\theta_r g_r(b\Delta)\right)\exp\left(\sum_{j=1}^{J}\gamma_j n_{k,b-j}\right) \qquad (4.10)$$

for trials $k = 1, ..., K$, which defines our likelihood function described by a GLM (Brillinger, 1988; McCullagh and Nelder, 1989; Kass and Ventura, 2001; Truccolo et al. 2005; Czanner et al., 2008). Modeling the CIF as a product of exponentials ensures that the firing rate is always strictly positive.

3.3   Empirical Mode Decomposition methodology

Empirical mode decomposition is a recently technique for analyzing non-stationary signals. It was developed for oceanographic studies in 1998 (Huang et al., 1998), but was quickly adapted for the study of biological signals. Unlike short time Fourier analysis or wavelet analysis, no underlying window or wavelet shape is assumed, and the modes of the decompositions emerge from the underlying temporal patterns in the signal at different timescales. A further advantage of this decomposition is its sparseness. Consider the decomposition of a linear chirp superimposed on a triangle wave (Figure 22A). Figure 22C through E shows the first three modes generated by performing EMD on the signal in Figure 22A. The largest modes closely resemble the linear chirp (Figure 22C) and the second mode (Figure 22D) the triangle wave. The third mode (Figure 22E) is much smaller than the first two and contains the residue. Therefore, this decomposition extracts the two dominant timescales in the signal in two modes to virtually perfectly reproduce the signal (Figure 22A). Short time Fourier analysis or wavelet transform decomposition of the signal would have necessitated many more modes since the underlying shapes are not periodic and do not resemble sinusoids or common mother wavelets. Modes derived using EMD are also locally orthogonal (Huang et al., 1998). This property means that neighboring components can carry the same frequency at different times; but, locally, any two components are orthogonal to one another (Huang et a.,

1998). However, orthogonality of the basis functions is not a necessary condition for our GLM framework.
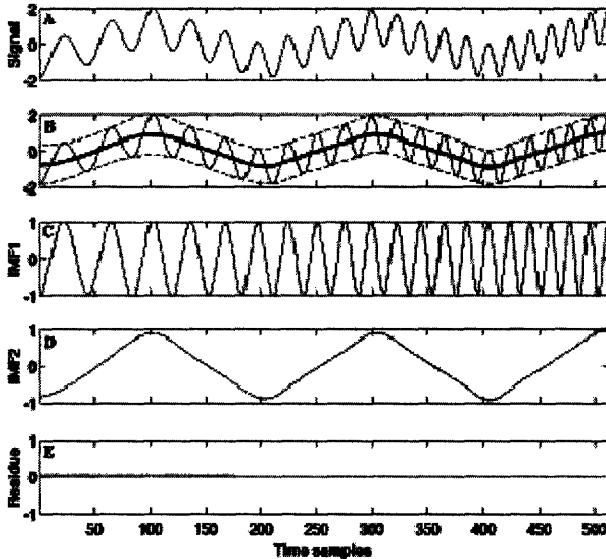

Figure 22:   Example of empirical mode decomposition from Liang et al. (2005). A. Signal is composed of a chirp superimposed on a triangular waveform. B. Reproduction of original signal (thin black line), the upper and lower envelopes (dotted lines) and point-by-point mean (thick black lines). C-E. Three modes from the signal. The top (C) corresponds to the chirp and the middle (D) corresponds to the triangular waveform, with the residue (E) represented as the third mode.

EMD is typically performed using a sifting procedure to iteratively extract modes from the underlying signal. By definition, an intrinsic mode function (IMF) has the properties that it is sufficiently narrowband (the number of extrema and zero crossings can at most differ by one) and the point by point average is zero. The sifting procedure first uses two splines to outline the upper and lower envelope of the signal, subtracts the point-by-point mean of the two envelopes and checks whether the resulting waveform satisfies the criteria for a mode. If it does not, the iterative procedure continues by subtracting the mean from the resulting signal, until the resulting waveform satisfies the mode criteria. Once a mode has been identified, the mode waveform is subtracted from the original signal and the sifting process continues to extract further modes until the stopping criteria is met. We use a normalized variance stopping criterion. The variance of the original signal is calculated and the variance of the residual signal is recalculated after every mode has been extracted. The sifting continues until the residual variance reaches a fraction of the original signal variance, where the fraction is based on the Crarmer-Rao bound. Typical values for the normalized variance stopping criterion are 10%. If the time-series $x(t)$ is decomposed into $j = 1: N$ IMFs, $c_j(t)$, then the relationship between the

143

signal, the IMFs and the residue, $r_N(t)$, follows this relationship:

$$x(t) = \sum_{j=1}^{N} c_j(t) + r_N(t).$$
(4.11)

Each IMF has a distinct timescale with the lowest-order mode having the shortest timescale and the highest-order mode having the longest (Huang et al., 1998).

3.4   Goodness-of-fit criteria and model order selection

3.4.1   Kolmogorov-Smirnov plot

Standard metrics of goodness-of-fit for continuous valued models, such as sum of squared errors, cannot be applied to point process data. An alternative method, utilizing the time-rescaling theorem (Ogata, 1988; Brown et al., 2001; Brown et al., 2002; Truccolo, 2005) has been specifically designed to conduct goodness-of-fit assessments on point process data. Once a parametric form of the CIF has been estimated, the rescaled interspike interval times can be computed as

$$z_{k,m} = 1 - \exp\left(-\int_{u_{k,m-1}}^{u_{k,m}} \lambda_k(u|\widehat{\varPsi})du\right),$$
(4.12)

where $u_{k,m}$ is the occurrence time of spike m at trial k for $m = 1, \dots, M_k$ and M_k is the total number of spikes in trial k. We assume that $u_{v,k,0} = 0$. We represent the ML estimates from the modeled CIF as $\widehat{\varPsi} = (\hat{\theta}, \hat{\gamma})$. The time-rescaling theorem states that, if the estimated CIF approximates the actual CIF well, the rescaled intervals $z_{k,m}$ will be independent and uniformly distributed on the interval [0,1). Therefore, a comparison between a uniform distribution and the rescaled intervals can assess how well the model corresponds to the true CIF. To assess the uniformity of the transformation, the rescaled intervals $z_{k,m}$ are ordered from smallest to largest $(z_{k,m}^*)$ and plotted against the values of a uniform cumulative density distribution defined as

$c_{m^*} = \frac{m^* - \frac{1}{2}}{M^*}$ for spike times $m^* = 1, \dots, M^*$ where M^* is the total number of interspike intervals. This is known as a Kolmogorov-Smirnov (KS) plot (Johnson & Kotz, 1970; Brown et al, 2002). If

the model is correct, the ordered rescaled intervals will lie along a 45-degree line. Confidence

bounds are calculated according to the KS statistic, and for medium to large data sizes, can be

well approximated by $c_{m^*} \pm \frac{1.36}{\sqrt{M^*}}$ for 95% certainty (Johnson and Kotz, 1970).

3.4.2 Autocorrelation function

To assess independence of the rescaled intervals, the spike times can be transformed

to a Gaussian distribution as $z_{k,m}^G = \phi^{-1}(z_{k,m})$, where $\phi(\cdot)$ is a cumulative distribution function

of a Gaussian random variable with zero mean and unit variance. If the Gaussian rescaled

times $z_{k,m}^G$ are uncorrelated up to some lag, then they are also independent up to that lag, since

lack of correlation implies independence for a Gaussian random variable. We assess the

correlation of the rescaled times $z_{k,m}^G$ by plotting the autocorrelation function and its confidence

intervals (Box et al., 1994). If the autocorrelation function remains generally within the

confidence bounds, then we can be confident that the original times are also independent. The

confidence bounds are calculated using Monte Carlo procedures with simulated draws from

identically distributed standard Gaussian random variables.

3.4.3 Model order selection using Akaike Information Criterion (AIC)

After fitting several models to the data we need to select the most suitable model. To do

so, we use the Akaike information Criterion (AIC) (Akaike, 1973) which has the form of:

$$AIC(p) = -2logPr(N_{1:K}|\widehat{\Psi}_p) + 2p \qquad (4.13)$$

where p is the total number of parameters in the model, $\widehat{\Psi}_p$ are the estimates of the model

parameters of order p, and $logPr(N_{1:K}|\widehat{\Psi}_p)$ is the log likelihood of the model as given in (4.5)

Models with a smaller AIC are generally considered superior in capturing more information in

the data than models with a larger AIC. The log likelihood decreases with the addition of each

parameter, however, the AIC decreases only when the decrease in the log likelihood is greater

than the addition of the extra parameter.

3.5 Decoding formulation

The discriminatability of a set of stimuli based on the response of a neuron can be assessed with decoding analysis. In decoding analysis, the maximum likelihood coefficients $\widehat{\Psi} = (\hat{\theta}, \hat{\gamma})$ are used to find the likelihood that the new data originated from the stimulus. We use decoding analysis on the set of responses to M_s songs with K_s trials for each song, s. To perform the analysis, we fit the GLM model to $K_s - 1$ of K_s trials for each of the M_s songs and obtain the ML coefficients $\widehat{\Psi} = (\hat{\theta}, \hat{\gamma})$. Using the remaining single trial from each of the M_s songs that was not used to obtain estimates for the ML coefficients, we compute the likelihood that the spike train in response to song m_s originated from a response to song m_r as:

$$L\left(s_{m_s}\middle|n_{m_r}\right) = \prod_{b=1}^{B}\left[\lambda\left(b\Delta\middle|H_b^{m_r}, \widehat{\Psi}_{m_s}\right)\Delta\right]^{n_{b_{m_r}}}\left[1 - \lambda\left(b\Delta\middle|H_b^{m_r}, \widehat{\Psi}_{m_s}\right)\Delta\right]^{1-n_{b_{m_r}}}. \qquad (4.14)$$

Next, we aggregate the likelihood over all M_s songs given spike train n_{m_r}:

$$L^{Total}\left(s_{1:M_s}\middle|n_{m_r}\right) = \left[L\left(s_1\middle|n_{m_r}\right) \quad L\left(s_2\middle|n_{m_r}\right) \quad \cdots \quad L\left(s_{M_s}\middle|n_{m_r}\right)\right], \qquad (4.15)$$

and take the maximum over the likelihood over the M_s songs as the most likely song from which spike train n_{m_r} originated:

$$L^{Total}_{MAX}\left(s_{1:M_s}\middle|n_{m_r}\right) = \max\left(L^{Total}\left(s_{1:M_s}\middle|n_{m_r}\right)\right). \qquad (4.16)$$

We then repeat this procedure for every other left out trial and for all songs.

3.6 Neural data description

The neural data came from the analysis of two earlier studies. Data were recorded in response to a spoken sentence in the mammalian auditory nerve and cochlear nucleus as well as to conspecific songs in the avian field L.

Mammalian responses:

Single unit responses were recorded from auditory nerve (AN) fibers and single-units from the cochlear nucleus (CN) from normal hearing cats anesthetized with Dial in urethane, as

146

described in detail previously (Cariani and Delgutte, 1996). Acoustic stimuli were delivered using a calibrated closed acoustic assembly, and were presented monaurally. The characteristic frequency (CF) of each unit was measured using threshold tuning curves (Kiang et al, 1970). For auditory nerve fibers, the CFs ranged from 121 Hz to 28.0 kHz (median 1.72 kHz) and the number of trials of stimulus presentations ranged from 19 to 39. For cochlear nucleus neurons, the CF range was 239 Hz to 32.7 kHz (median 6.90 kHz) and the number of trials range was 19 to 100. The neural recordings analyzed here include responses to a speech utterance sampled at 20 kHz. The utterance was the IEEE sentence "Wood is best for making toys and blocks" verbalized by a male speaker (Delgutte et al., 1998). The peak sound pressure level for the speech was 60 dB SPL.

Avian responses:

Single unit responses were recorded from field L of male zebra finches anesthetized with intramuscular injections of urethane. Recordings were performed in a double-walled sound-attenuated chamber with the animal facing the loudspeaker, located 20 cm away from the beak. The animal was elevated to the level of the speaker cone. Conspecific song stimuli were played at a peak intensity of 75 dB SPL and the current dataset consists of responses to 10 trials of 20 songs from one single recording site. Responses were sorted into single units using spike shape (Sen et al. 2001). Electrode placement was verified with electrolytic lesions and histological markers within the boundaries of field L were assessed. The songs were undirected and conspecific zebra finch songs recorded in a sound-attenuated chamber. The songs were sampled at 32 kHz and bandpass filtered to retain frequencies between 250 and 8000 kHz (Narayan et al. 2006).

4. RESULTS

4.1 Single-filter point process model of auditory nerve data

We begin our analysis of auditory single unit responses to natural stimuli by modeling a response to a sentence in the auditory nerve. We select the auditory nerve, the primary nucleus in the auditory system, to present a simplified version of our model decomposition. In later sections, we discuss the general version of our GLM model where the CIF of the neural response is modeled as $\lambda_k\left(b\Delta|\theta,\gamma,H_{k,b}\right) = \exp\left(\sum_{r=1}^{R}\theta_r g_r(b\Delta)\right)\exp\left(\sum_{j=1}^{J}\gamma_j n_{k,b-j}\right)$, for an arbitrary number of basis functions $g_r(b\Delta)$. However, early work has shown that each auditory nerve fiber functions as bandpass filter tuned to the fiber's characteristic frequency (CF) in encoding the sound stimulus and each auditory nerve fiber is connected to a single hair cell (Geisler, 1998). Therefore, we begin the analysis by constraining the CIF to include only one basis function $g_1(b\Delta)$.

Following the view that the auditory nerve fibers act as bandpass filters, typical models of the auditory periphery include a two stage implementation (e.g. Geisler and Greenberg, 1986). The first stage models the filtering properties of the auditory nerve fiber, largely derived from the fiber's location along the basilar membrane (Geisler, 1998). Recent recordings in the auditory periphery in animals have developed models for the filter transfer functions. To represent the filtering stage, we use a gammatone filter based on auditory nerve data in cat (Carney, 1993) that represents the fiber bandpass filter characteristics as a function of the fiber's CF, the frequency of greatest sensitivity at the lowest stimulus level. Therefore, our single basis function in this first simple model will be the impulse response of the gammatone filter appropriate for a particular auditory nerve fiber.

The second stage represents the hair cell and its synapse with an afferent neuron. To a first approximation, this transformation can be quantified as half-wave rectification followed by low-pass filtering. Half-wave rectification arises from the inner hair cell voltage versus

displacement characteristic (Hudspeth and Corey, 1977) and from the transformation of receptor potential into spikes, as spike probabilities are strictly positive. The inner hair cell voltage versus displacement characteristic is such that for small displacements, the input/output function is relatively linear, but becomes increasingly compressive in both directions for larger displacements. Since the operating point is located much closer to the compressive region for negative displacements, approximate half-wave rectification results. The lowpass filtering is attributed to the inner hair cell membrane time constant and the jitter in the synapses between the hair cell and the auditory nerve. These models have qualitatively accounted for the responses seen at the auditory nerve (e.g. Geisler and Greenberg, 1986). In order to utilize this model in the context of our framework, we add one final stage to the two-stage model of the auditory periphery. As described in Methods, our neural responses are captured at a resolution, Δ, of 1 ms following the typical refractory period of neurons. Therefore, we decimate the output of the lowpass filter by a factor of $M = 20$, where M is the ratio between the original sampling rate of the signal and 1 kHz. Figure 23 shows the lowpass filter used to partially model the functioning of the inner hair cell membrane time constant and synapse jitter combined with the lowpass filter used as part of the downsampling procedure. Note that this model captures the envelope of the response.
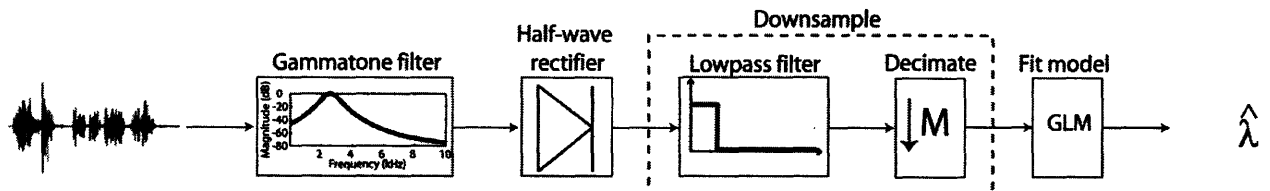

Figure 23: Illustration of methodology for neural response model estimation using gammatone filters. The speech signal (gray) sampled at 20 kHz is filtered using a gammatone filter with the center frequency equal to the auditory nerve fiber characteristic frequency (CF) and halfwidth equal to $\frac{1}{2\pi\tau}$, where the filter time constant is equal to $\tau = C_0 exp\left(-\frac{x}{S_0}\right) + C_1 exp\left(-\frac{x}{S_1}\right)$ in seconds, $C_0 = .0011$ seconds, $C_1 = .0011$ seconds, $S_0 = 6.0$ mm, $S_1 = 2.2$ mm and $x = 11.9 log_{10}(.8 + \frac{CF}{456})$. This methodology uses the Carney auditory filter for auditory nerve data at 75 dB (Carney, 1993). Next, the output of the filter is half-wave rectified and downsampled to a rate of 1 kHz. The lowpass filter in the downsampler is an eighth order Chebyshev Type I filter with cutoff frequency of 400 Hz. Lastly, a GLM model is fit to the output of the downsampler and past spiking history covariates to estimate the conditional intensity function.

We implemented this class of models in the generalized linear model (GLM) framework to show that our framework can describe the transformation of the speech sound to neural responses. We illustrate this model using the auditory nerve response in the dot raster shown in Figure 24A for 30 trials of the speech stimulus shown in gray. This fiber has a fairly high background firing rate, as shown by the response during brief silences in the speech stimulus, and fires vigorously for several of the phonemes. We follow the two stage model of the auditory periphery to construct a model of the neural response from the sound stimulus. We call this model the GLM-gammatone model for auditory nerve responses in the point process GLM framework.

As described in methods, the GLM-gammatone model consists of the portion of the response attributable to the stimulus, dependent on the stimulus decomposition, $g(b\Delta)$, and the portion of the response attributable to the effect of the past spiking history of the neuron. To model the stimulus-effect component, we use the output of the two-stage filter in Figure 23 shifted by some number of small time intervals, ranging from $d = 1, \dots, D$ ms. These time-shifts represent both axonal delay and filtering occurring at the synapse. To represent the past spiking history, we model the response for the past $j = 1, \dots, J$ ms. Therefore, our GLM-gammatone model for the auditory periphery is formulated as:

$$\lambda_k(b\Delta|\theta, \gamma, H_{k,b}) = \exp\left(d_0 + \sum_{d=1}^{D} \theta_d g_d(b\Delta)\right) \exp\left(\sum_{j=1}^{J} \gamma_j n_{k,b-j}\right), \qquad (4.17)$$

where d_0 represents a baseline response often included in models of the auditory periphery (e.g. Delgutte et al. 1998). For our simple models of the auditory peripheral response, we examined delays up to $D = 20$ ms and past spiking histories of up to $J = 100$ ms.
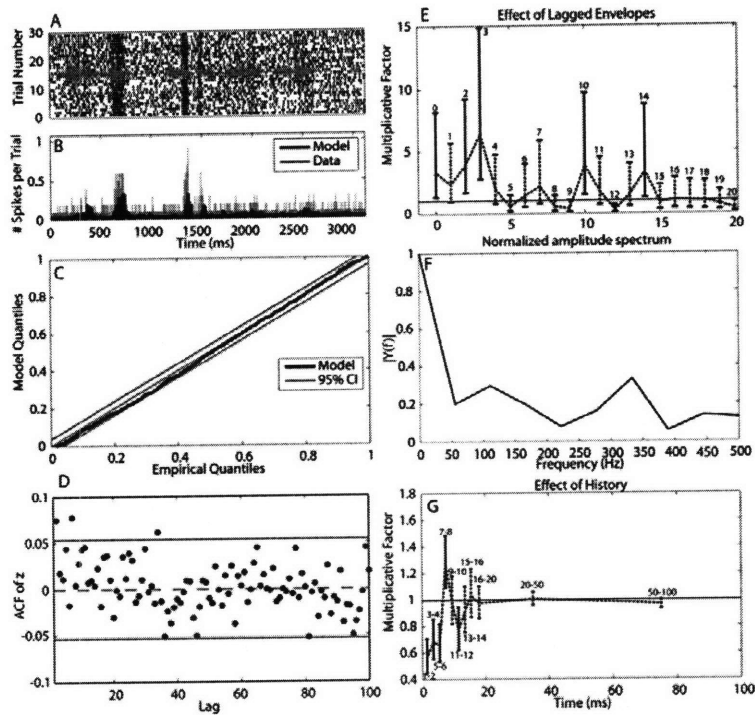
150

Figure 24: GLM-gammatone model and goodness-of-fit A. Dot raster plot of 30 trials of responses to speech stimulus (gray). **B.** Peristimulus time histogram (PSTH) of response in A (gray) and sum of CIF across trial (black). **C.** Kolmogorov-Smirnov (KS) plot showing model goodness-of-fit. **D.** Autocorrelation function of Gaussian rescaled intervals. **E.** Exponentiated coefficients describing the effect of lagged stimulus envelope outputs of the gammatone filter at the neuron's characteristic frequency (CF=2.579 kHz). Coefficients and 95% confidence intervals are shown. Solid confidence intervals represent coefficient estimates significant at a 0.05 level of significance, while dotted confidence intervals represent coefficient estimates that are not significant at a 0.05 level of significance. **F. Normalized amplitude spectrum of lagged envelope effects. G.** Exponentiated coefficients describing the effect of past stimulus history on current spiking and 95% confidence intervals.

Figure 24E and G show the model parameter estimates for the GLM-gammatone model of the auditory periphery for one example response in Figure 24A. First, observe the stimulus component of the response (Figure 24E). The model parameter estimates indicate that the unit responds for delays of zero (indicating some rate-based baseline response), two, three, ten and fourteen ms after the stimulus onset following the pattern of the two-stage model of the periphery. Other shifts of the covariates are not significant at a 95% level. The normalized amplitude of the spectrum is shown in Figure 24F, highlighting that the parameter estimates indicate stronger responses to envelope frequencies of about 100 Hz, 325Hz and DC. Second, note the shape of the history response dynamics (Figure 24G). The dynamics are shown as a multiplicative factor, such that a value of 1 indicates that previous spikes do not change the instantaneous firing rate, >1 indicates an enhancement of the response and <1 a suppression of the response. For the first six ms after a previous spike, our model estimates indicate that the neuron exhibits refractory behavior, such that the response is decreased over what would be

151

expected by the stimulus effect alone. From seven to eight ms after a previous spike, the cell is slightly more likely to emit a spike and for eleven to twelve ms after a previous spike, the cell is again slightly refractory. Spiking history with a longer duration than twelve ms does not have a significant effect on the response of the cell.

To examine the goodness-of-fit of our model, we show the Kolmogorov-Smirnov plot that indicates the extent to which the CIF captures the stochastic nature of the response on a spike-by-spike basis (Figure 24C). As discussed in Methods, because the plot lies along a 45-degree line and within the 95% confidence bounds for the Kolmogorov-Smirnov statistic, we conclude that the model provides an excellent fit to the data. The excellent model fit is also highlighted by the autocorrelation of the rescaled intervals (Figure 24D) which shows no significant correlations up to lag 100. These measure of goodness-of-fit are appropriate for point-process models that capture both the rate-based and timing aspects of the response. On the contrary, comparisons between the actual (Figure 24B gray) and estimated (Figure 24B black) PSTH are inappropriate for goodness-of-fit assessment of these dynamic models because they average over the history of the response (see Discussion).

The goodness-of-fit and interpretability of the two-stage peripheral model, extended within our framework as a point process GLM formulation with past spiking history indicates that the two-stage peripheral model can be well represented in our framework. We replicated this analysis for our entire dataset of auditory nerve responses to show the appropriateness for utilizing our framework for this data. To summarize the results, we show goodness-of-fit assessments for the dataset. One method for assessing the goodness-of-fit of the GLM-Gammatone method is to examine the Kolmogorov-Smirnov plots for the dataset. The KS plot describes the goodness of fit *relative* to other models for the same neurons because the size of the confidence bounds of the KS statistic ($\frac{1.36}{\sqrt{M}}$ for a 95% confidence interval where M is the total number of spikes) depends on the total number of spikes in the response. Therefore, one

cannot compare the KS plot of models for two different neurons by plotting the results on the same KS plot.

However, one derived measure of the goodness-of-fit from a KS plot is the KS statistic, which measures the largest deviation from the diagonal on a KS plot. If we normalize the KS statistic relative to one side of the 95% confidence interval, we obtain a normalized KS statistic. For example, if the largest deviation from the diagonal is exactly the upper bound of the 95% confidence interval, the value of the relative KS statistic is 1. By examining the range of the relative KS statistics in the dataset, we visualize the overall goodness of fit of the data. However, it is important to remember that the KS plot is a very conservative measure of the uniformity of the rescaled intervals. It is a conservative measure because the KS plot examines the cumulative distribution of interspike intervals. For instance, if only one interval deviates significantly from the expected distribution, the plot will likely fall outside the 95% CI for a large portion of the KS plot, until another interspike interval falls outside the confidence bounds in the other direction. It is therefore much more informative to examine the shape of the KS plots for a close to linear (as opposed to a bowed) plot, since this type of plot generally indicates a rather appropriate model fit despite exceeding the confidence bounds.

Because the dataset is large in this example, we present the assessment of the goodness-of-fit by KS plots by showing the best and worst KS plots as well as the range of the KS statistics. The best KS plot for the auditory nerve dataset (Figure 25A) is very close to linear and does not deviate at all from the 95% confidence interval (relative KS statistic < 1). Notice that the firing rate of this unit is fairly low, and consequently, the confidence bounds of the KS statistic are fairly large. On the contrary, the response with the worst KS statistic (Figure 25B) has an extremely high firing rate, so high that visually, the high firing rate looks virtually sustained at the magnification of the dot raster shown in the figure. The resulting confidence bounds are therefore very small (and much smaller than that of the best KS plot shown), resulting in a larger relative KS statistic (approximately 6.5). However, notice that the shape of

153

the KS plot is still fairly linear, although the plot does deviate from linearity in at least two different ranges of the intervals. The shape of the KS plot and the KS statistic indicates that the model would do a fair job in predicting the probability of each spike. Overall, the histograms of the KS plots show the largest deviations of the KS statistics relative to the confidence bounds. All but 4 responses (of 83) fall within three times the size of the 95% confidence interval. The responses with the largest KS statistics are those that have the largest firing rates. We conclude that the overall goodness-of-fit of the models as indicated by the KS plots is satisfactory for the entire dataset in predicting the probability of the consequent spike on a spike by spike basis. Therefore, the GLM-gammatone framework is an appropriate one for representing auditory peripheral data using a two-stage model within a point process GLM framework.
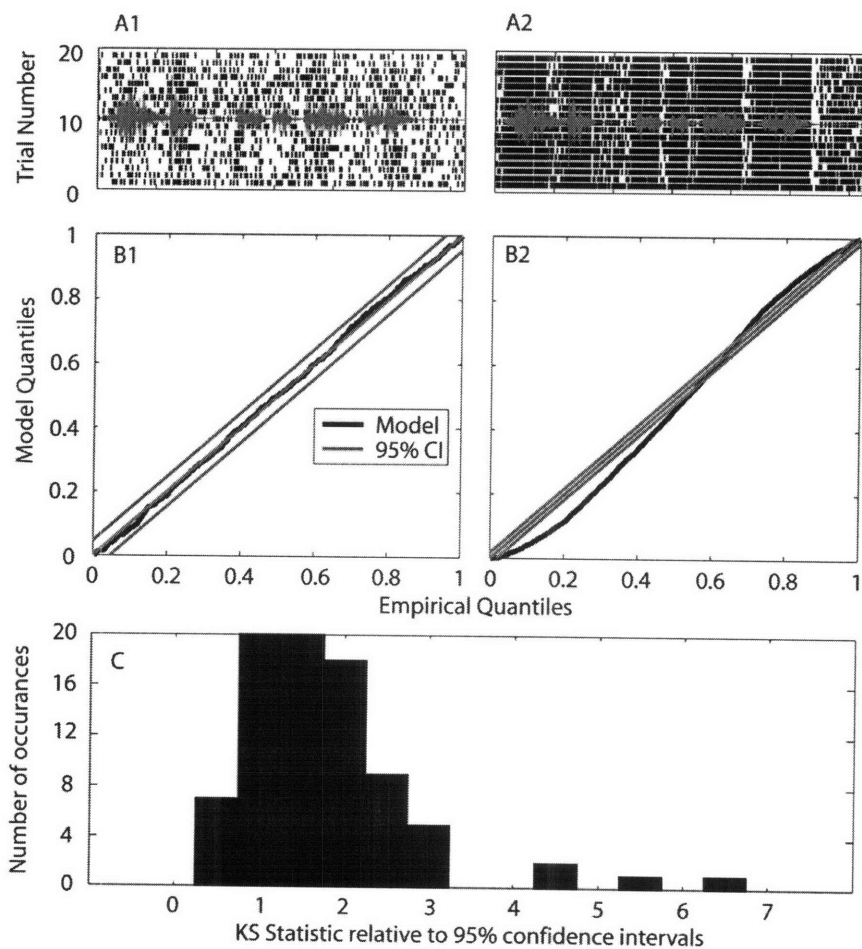

Figure 25: Range of goodness-of-fit of GLM-gammatone model for auditory periphery in point process GLM framework. **A1:** Dot raster for response of neuron with best KS plot in dataset. Sentence stimulus shown in gray. **B1:** Best KS plot in dataset. Gray lines indicate 95% confidence intervals for the KS statistic. **A2:** Dot raster of response with the worse KS plot in dataset. **B2:** Worst KS plot in dataset. **C:** Histogram of KS statistic relative to 95% confidence interval for entire auditory nerve dataset. A value of 1 indicates that the KS statistic (greatest deviation of KS plot from diagonal) is just within the 95% confidence bounds of the KS statistic for that response.

4.2   Point process models using Empirical Mode Decomposition

4.2.1   Models of responses in the auditory nerve and cochlear nucleus in response to a spoken sentence

To model responses in auditory centers above the auditory nerve, the model of cochlear processing as a single bandpass cochlear filter may not be sufficient. In the next stage of processing, the cochlear nucleus, axons from several auditory nerve neurons can converge on one cell (Cant, 1992). The convergence becomes even more dense in higher auditory centers, and can involve efferent projections from upstream neurons as well as projections from multiple downstream nuclei (Geisler, 1998). Since the neural inputs at a specific nucleus are often not well known, an alternative approach looks at the aspects of the stimulus that can most closely predict the neural response of a specific cell. In addition, because inputs may involve innervations from different nuclei with different axonal lengths and computational durations, one important aspect of the analysis involves estimating the delay of a particular stimulus aspect relative to the cellular response.

In this section, we utilize EMD decomposition to capture the different time-scales of the acoustic stimuli because of the sparseness of the decomposition and the applicability to natural stimuli that are non-stationary. We use the EMD decomposition to describe the stimulus covariates while the neuron's own spiking history captures additional factors that may influence the future response. In this section, we illustrate the use of EMD decompositions in the GLM framework to model the responses of mammalian neurons in the auditory nerve and cochlear nucleus to the same spoken sentence.

Figure 26 illustrates the EMD of the sentence that is used as the acoustic stimulus for the responses in the auditory nerve and the cochlear nucleus neurons. The time-domain representation of the sentence, originally sampled at 20 kHz is shown in Figure 26A. In this plot, the times during which the amplitude decreases correspond to lip closures and times with large signal amplitudes correspond to phonemes, particularly those that are voiced, such as /l/

in 'is' at .7 sec and /A/ in 'blocks' at 2.6 sec. Figure 26B shows the spectrogram of the sentence along with IPA markings indicating the locations of each phoneme.
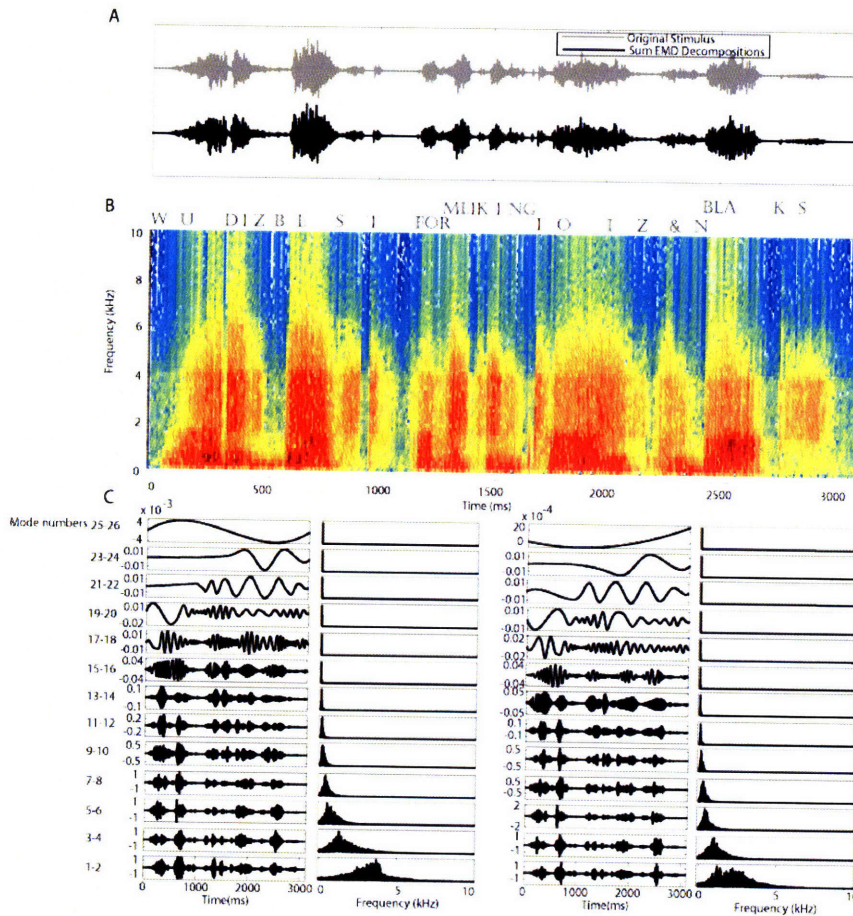

Figure 26: Empirical mode decomposition of sentence "Wood is best for making toys and blocks." A. Original sentence sampled at 20 kHz (gray) and sum of 26 EMD intrinsic mode functions (IMFs). B. Spectrogram of sentence with IPA markings. The /R/ in 'for' and /M/ in 'making' are co-articulated, as are the /NG/ in 'making' and /T/ in 'toys' and the /N/ in 'and' and /B/ in 'blocks'. The /D/ in 'and' is not articulated as the two voiced consonants /N/ in 'and' and /B/ in 'blocks' blend together. C. Twenty-six IMFs and their frequency spectrum. For each mode, the time domain mode sampled at 20 kHz is shown on the left and the magnitude of the fast Fourier transform is shown on the right. The spectra are not to scale.

This sentence, shown in Figure 26A gray, is decomposed, using EMD, until the size of the residue meets the stopping criteria as described in Methods. For this sentence, 26 modes are sufficient to meet the stopping criteria, and they are plotted in Figure 26C. The modes with the largest amplitudes are those of the lowest-order numbers. The order of the decomposition process is such that the lowest-order modes are extracted first (mode 1) and the highest-order last (mode 26). These 26 modes very adequately represent the original song stimulus, and their sum is shown in Figure 26A (black). Notice that it is impossible to visually discriminate between the original stimulus and the sum of the decompositions by eye.

Just as each mode captures a different time-scale represented in the stimulus, the spectrum (Figure 26C) of each mode differs. For example, the high-order modes represent very

low frequency components, while the low-order modes represent more temporally complex components whose spectrum spans a larger frequency range over the entire duration of the song. Note that we show the spectrum, averaged over the song duration, for each mode. Actually, the instantaneous frequency spectrum of each mode is quite narrowband. Huang et al. (1998) sees the instantaneous narrowband nature of the EMD as a major asset to the decomposition, and illustrates this aspect by extracting these components using the Hilbert transform.

In our model, we examine the effect of the EMD modes and the neuron's own past spiking history on the current propensity to generate spikes in response to the sentence stimulus. However, as we noted in the description of the GLM-gammatone model in the previous section, the model needs to include some peripheral processing of the EMD covariates in order to represent the peripheral processing of the inner ear. We therefore half-wave-rectify and downsample each mode, using the same procedure as described in Figure 23 with the GLM-gammatone filter model. To represent the filtering and axonal delays, we include time shifts of the rectified envelopes as the model covariates in addition to the neuronal spiking history. The full neuronal model used in this section is that shown in Eq. (4.10) , where $g_r(b\Delta)$ are the half-wave rectified and downsampled modes of the EMD with delays.

4.2.2 Responses in auditory nerve and cochlear nucleus

Figure 27 shows an example response in the auditory nerve (CF=487 Hz) to the sentence. Figure 27A shows the stimulus overlaid on a dot raster plot of responses to 20 stimulus presentations. In the dot raster, each black dot represents a spike. As evidenced from the dot raster plot and also from the PSTH of the response (Figure 27B), the cell's response increases during the vowels /U/ in 'wood', /OI/ in 'toys' and /A/ in 'blocks' (see spectrogram in Figure 26B). This observations stems from the relatively low-frequency nature of the unit, since its characteristic frequency is 487 Hz, and the vowels /U/, /OI/ and /A/ have frequency content in that region. The first formants of the vowels /U/, /O/, /I/ and /A/ are approximately centered at

157

350 Hz, 500 Hz, 350 Hz and 700 Hz, respectively (Stevens, 2000), all most likely within the
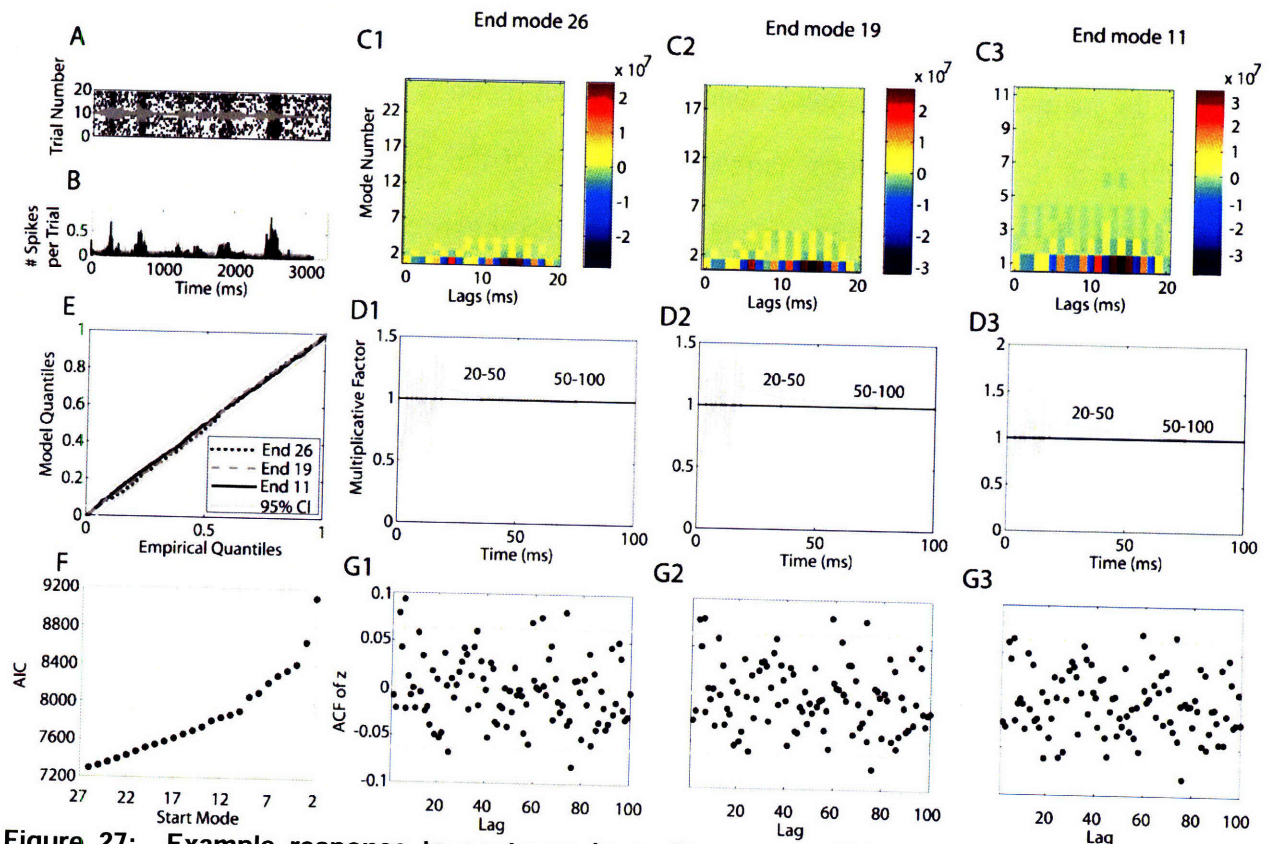
cell's filter response (Carney, 1993).

Figure 27: Example response to sentence in auditory nerve (CF=487 Hz). A. Dot raster of responses to 20 trials of sentence (shown in dark gray). B. PSTH (light gray) and sum of conditional intensity function over trials (black) for response shown in A using full model with all 26 modes (C1 & D1). C1-3. Estimates of lagged envelopes of mode parameters for model ending at mode 26 (26 models total), at mode 19 (19 modes total) and mode 11 (11 modes total). Lags used range from 0 to 20 ms using 1 ms bins. D1-3. Estimates of multiplicative history components for models ending at mode 26 (D1), at mode 19 (D2) and at mode 11 (D3). History effect parameters are used for the following history bins: 1-2, 3-4, 5-6, 7-8, 9-10, 11-12, 13-14, 15-16, 16-20, 20-50, 50-100 ms. E. KS plot for each of the three models. The 95% confidence intervals for the KS statistic are shown in light gray. F. AIC for all models starting with only 2 modes (1 and 2) and with each additional mode through model with all 26 modes. G1-3: Autocorrelation function of Gaussian rescaled intervals for each of 3 models. Light gray lines show the 95% confidence bounds.

Figure 27C and D show plots of the covariates of the model, the lagged rectified

envelopes and the effect of the past spiking history. First observe the effect of the lagged

rectified envelopes on the response in Figure 27C1. The size of the parameters is largest for

the lowest-order modes (1-3), while the higher-order modes have parameter values close to

zero. The lowest-order modes also have the largest amplitudes, so the size of the parameters

is particularly strong indicator of a larger effect of these modes. This model is also implemented with normalized lagged rectified envelopes (not shown), with the relative sizes of these parameters unchanged by the normalization. We also examine models containing fewer of the high-order modes since the smaller magnitudes of their parameter estimates suggest that their contribution is less important to the model description. For instance, Figure 27C2 shows a model with the same history covariates as the full model in Figure 27C1, but with lagged rectified envelopes ranging from modes 1 to mode 19. Likewise, Figure 27C3 shows a model with lagged rectified envelopes ranging from modes 1 to 11. The pattern of the estimated parameter values of the lagged rectified envelopes is generally similar for all three models, with the largest contribution arising from the lowest numbered 2-3 modes.

The parameters describing the effect of the past spiking history for the three models are generally similar, although small differences do exist between the models. The full model (Figure 27D1) history parameter estimates indicate that the cell is refractory for the first 1-2 ms following a spike. However, for all stimulus histories more than 2 ms in the past, past spiking activity does not significantly influence current spiking, as indicated by the 95% confidence intervals of the parameter estimates including 1. Remember that a history effect of 1 indicates that the previous spiking does not affect current spiking. The history effect follows a similar pattern for the reduced model in Figure 27D2 (modes 1 through 19), with previous spiking history in the 1-2 ms range the only significant past spiking history modeled. However, the further reduced model in Figure 27D3 (modes 1 through 11) has an additional significant parameter for previous spikes 20-50 ms in the past. This gradual increase in stimulus parameters may reflect an omitted variable bias in the model (e.g. Greene, 2007). One should generally begin with a larger model and use the smallest model for which the model parameters do not deviate significantly from the parameter values in a larger model.

Whenever examining model parameters, it is also important to concurrently examine model goodness-of-fit. To do so, we employ two metrics, the Kolmogorov-Smirnov plot to

assess the uniformity of the rescaled intervals and the autocorrelation of the rescaled intervals to assess their independence (see Methods). The KS plots for the three models are shown in Figure 27E. For each of the three models, the KS plot falls completely within the 95% confidence intervals for the KS statistic, indicating that the model captures the stochastic nature of the response. The autocorrelation of the Gaussian-rescaled intervals also gives good feedback to this model. For each of the three models (Figure 27G1, G2 and G3), the rescaled intervals do not follow a discernable pattern of correlation with previous lags and at least 93% of the observations are within the 95% confidence bounds for this assessment. Since uncorrelated implies independence for Gaussian random variables, we can conclude that the rescaled intervals are also independent up to lag 100 (and uniform as indicated by the KS plot), and can therefore conclude that the goodness-of-fit for each of these models is adequate.

However, it is important to examine one more model assessment before concluding which model order to select. The Akaike Information Criterion (AIC) is a model selection criterion that trades off model complexity against how well the model fits the data (see Methods). The model complexity is determined by the number of parameters describing past spiking history and the number of parameters needed to describe the contribution of the lagged modes. AIC calculations for models with decreasing number of modes are shown in Figure 27F. For example, a model with end mode 2 only includes lagged rectified envelopes of modes 1 and 2, whereas a model with end mode 26 is the full model whose parameter estimates are shown in Figure 27C1 and Figure 27D1. We generally select the model order for which the AIC exhibits concavity and a decrease in the size of the decrease with the addition of each parameter (e.g. Czanner et al., 2008). For this example, there is a decrease in the rate of decrease at the model for modes 1 through 10, however, one could also make a case for selecting the full model with an AIC plot such as the one shown in Figure 27F. Therefore, it is important to consider all elements of the goodness-of-fit analysis (KS plot, autocorrelation and AIC) and the tolerance for omitted variable bias before selecting the appropriate model order for

this framework.

Neurophysiologists typically neglect the contribution of spiking history in their models of the neural response. More specifically, the goodness-of-fit of a model is typically gauged by the closeness is the estimated PSTH and the actual PSTH of the response. It is possible to compute an estimate of the PSTH in our case as well by averaging the CIF over trial (Figure 27B in gray), however, this approach averages over the temporal dynamics of the estimated CIF, the incorporation of which is a major advantage of our analysis. Therefore, we show this estimated PSTH for completeness, but do not consider it an appropriate assessment metric for point process models. Although our estimated PSTH is virtually indistinguishable from the actual PSTH in this case, such will only be the case when the temporal dynamics are relatively unimportant in the model. Otherwise, the KS plot and the autocorrelation serve as much more rigorous assessments of the goodness-of-fit of the model.

We also examined responses in the cochlear nucleus in the GLM-EMD framework using the full model and two reduced models as described for the auditory nerve example above. The response shown in the dot raster in Figure 28A has similar characteristics to those seen in the auditory nerve example. Particularly, the response increases for the voiced vowel phonemes /U/ in 'wood', /I/ in 'is', /OI/ in 'toys' and /A/ in 'blocks'. The PSTH of summed and normalized spike times is shown in Figure 28B (black) as well as the estimated PSTH using the GLM-EMD model computed by adding the conditional intensity function across trial and normalizing (gray).

Similarly to the response in the auditory nerve, the parameters describing the contribution of the lagged rectified envelopes of the EMD modes are largest for the first 2 modes. The lowest-order mode magnitudes are restricted to stimulus lags around 5 ms and around 15 ms, and this pattern is generally seen for each of the three models. The multiplicative history parameters for this example tend to show longer duration of significant refractoriness than responses in the periphery, for which the neuron only showed significant refractoriness for the first 1-2 ms following a spike (Figure 27D1, D2 and D3). Instead, this

161

cochlear nucleus response shows significant refractoriness for 1-2, 3-4, and 5-6 ms following a spike for all three models, for 7-8 ms for the full and larger of the reduced models (end mode 19) and for an additional 9-10 ms following a spike for the full model (Figure 28D1, D2 and D3). Again, these differences in parameter estimates between the three models can be attributed to omitted variable bias that should be carefully examined when eliminating modes from the model with respect to a tolerance level.
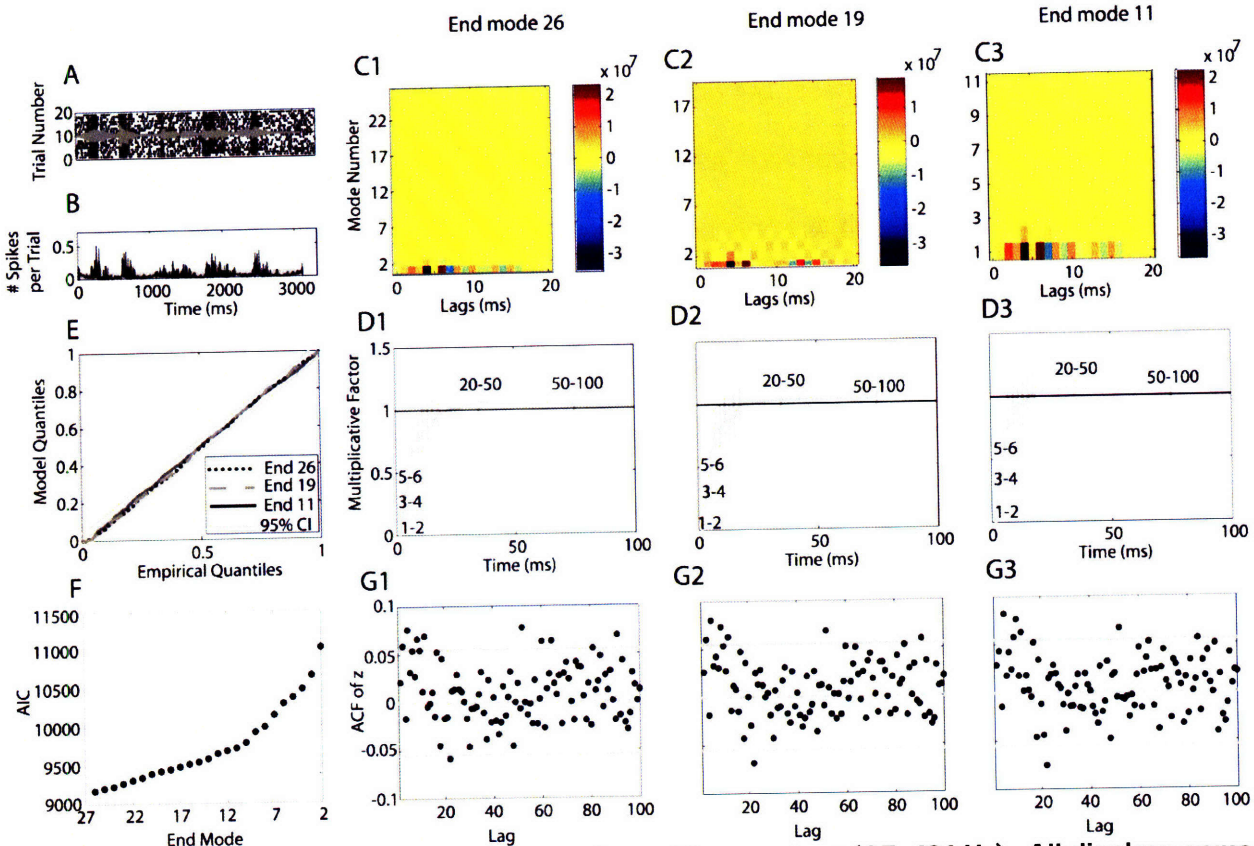

Figure 28: Example response to the sentence in cochlear nucleus (CF=494 Hz). All displays same as those shown in Figure 27.

The goodness-of-fit of each of these 3 models is highly adequate. The KS plots for the full and reduced models (Figure 28E) indicate that the rescaled intervals are all within 95% confidence intervals of the KS statistic, indicating that the rescaled intervals are uniform. The autocorrelation of the Gaussian-rescaled intervals do not show dependence up to lag 100 and very few points lie outside the 95% confidence intervals for the rescaled intervals, indicating that the rescaled intervals are also independent up to lag 100 (Figure 28G1, G2 and G3) . The

uniformity and independence of the rescaled intervals indicate a good model fit to the data for each of the 3 models. Lastly, the AIC can be used to evaluate the tradeoff between selecting a more reduced model, thereby reducing the number of parameter estimates, and obtaining a larger goodness-of-fit. The change in concavity of the AIC at start around end mode 11, indicates that a model with modes 1 through 11 may present an adequate tradeoff. However, the size of the model should be selected for each application relative to the tolerance in AIC.

We next examine the model parameter estimates for the entire AN and CN dataset. Beginning with the history parameter estimates, Figure 29A illustrates the affect of previous spiking history on current spiking across the AN and CN datasets for the model with 26 modes. Both datasets indicate that the spiking activity in the first 1-2 ms following a previous spike suppresses the response, with the propensity to spike reduced to approximately 60% for the AN dataset. Also in the AN, for the next several time bins following a previous spike (3-4, 5-6, 7-8, 11-12, 13-14 ms, 9-10 ms bin excluded) the affect on current spiking is still significantly suppressive at a 5% level of significance (as indicated by the notch surrounding the median value below 1), although the extent of the suppression is reduced. The suppression no longer plays a role in current spiking past 15 ms of previous spiking history. The non-parametric one-way analysis of variance Kruskal-Wallis test rejected the hypothesis that the samples at each time bin came from distributions with equal medians ($\chi^2(10,814) = 253, p \ll 0.001$). In this study, we find comparable, but slightly longer significant effects of previous spiking history than those found by Miller and Mark (1992), who find suppressive effects in two AN fibers until 10 ms after a previous spike, and indicate that the results "are typical for the majority of the auditory nerve fibers" in response to a vowel stimulus. The temporal progression of the affect of previous spiking history in the CN dataset is similar, with very strong suppression seen during the first 1-2 ms following a previous spike and with gradual diminishing of the strong suppression until 14 ms, and no significant suppression for 15 ms following a spike. The temporal progression is again significant as indicated by the Kruskal-Wallis test ($\chi^2(10,748) =$

163

193, $p \ll 0.001$). To the authors knowledge, the effect of previous spiking history in the CN has not been previously reported. However, the degree of initial suppression in the CN dataset is much greater than that seen in the AN dataset, with the spiking 1-2 ms following a previous spike reduced to approximately 20% of the original spiking propensity, compared to 60% in the AN. In addition, the distributions of the parameter estimates are wider in the CN. A Wilcoxon ranksum test rejected the hypothesis that at the AN and CN data come from distributions with equal medians for all parameter estimates up through 15-16 ms ($p < 0.05$) but not for parameter estimates beyond 16 ms. Therefore, the suppressive effect of previous spiking history differs for the AN and CN populations as a whole.
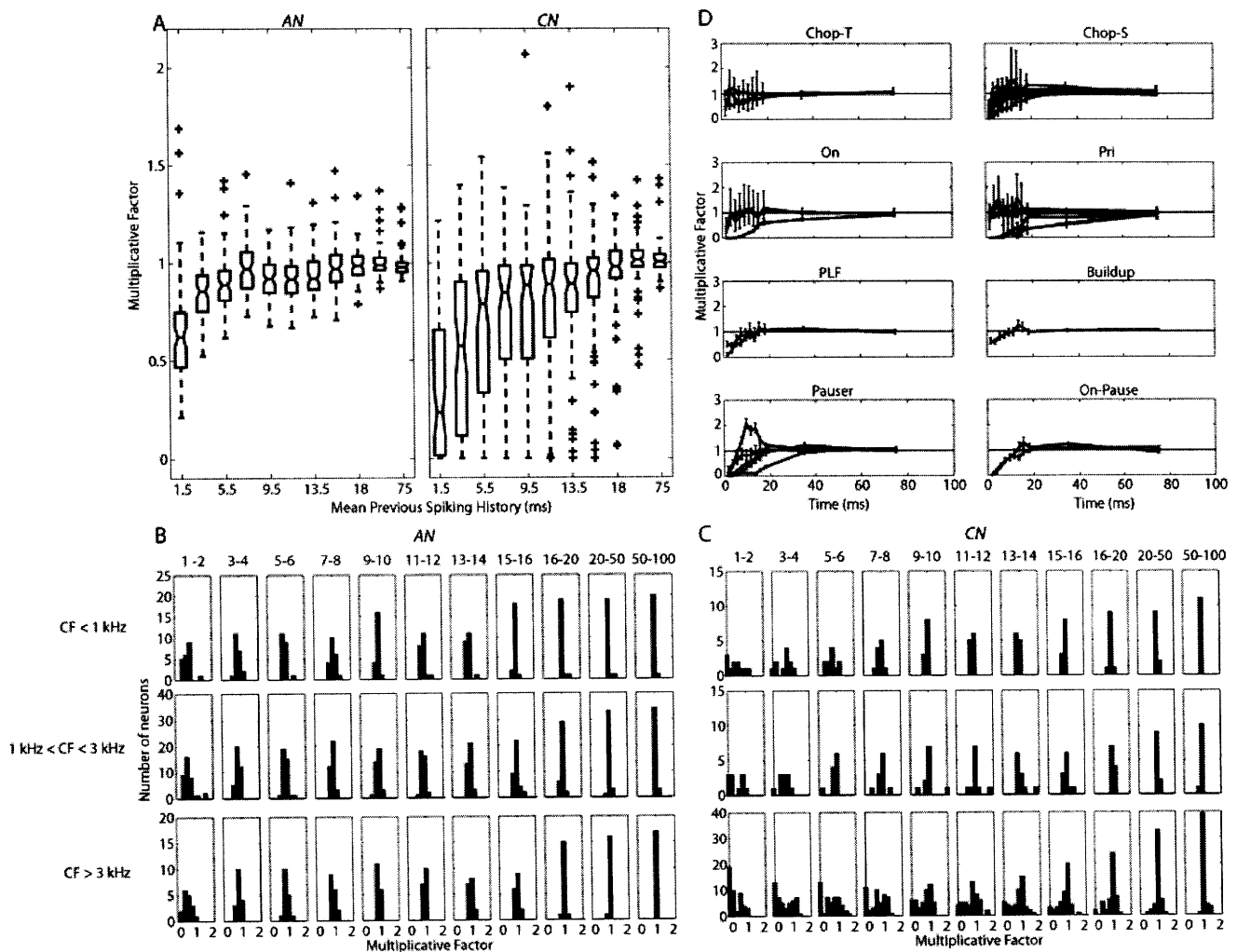

Figure 29: Parameters describing affect of previous spiking history for auditory nerve and cochlear nucleus data derived from the GLM-EMD model containing 26 modes. A. Boxplots of the

multiplicative history parameters for the entire AN and CN datasets. Boxplots outline the 25th and 75th quantiles, with the median indicated by the horizontal line. The whisker width is 1.5 times the interquartile range and the pluses indicate outliers. The notches illustrate a visual hypothesis test, and if the notches do not overlap, it can be concluded with 95% confidence that the true medians for each of the parameter values differ. History bins in the model are 1-2, 3-4, 5-6, 7-8, 9-10, 11-12, 13-14, 15-16, 16-20, 20-50 and 50-100 ms of previous spiking history. In this subplot, they are indicated by their mean. B. Histograms indicating the multiplicative factors describing the affect of previous spiking history in each modeled time bin as a function of the characteristic frequency of the neuron. Neurons are grouped into three groups: those with CF less than 1 kHz, with CF between 1 and 3 kHz and with CF greater than 3 kHz. C. Same as B for CN dataset. D. Effect of previous spiking history for eight groups of CN neurons in the dataset. Approximately 1/3 of neurons in the dataset were previously classified according to these 8 groups and are shown here. Thick black lines indicate the median parameter values and errorbars indicate 95% confidence intervals around the estimates. Neurons were classified by P. Cariani according to their firing patterns in response to short tone bursts (Pfeiffer, 1966; Bourk, 1976; Rhode and Smith, 1986; Blackburn and Sachs, 1989). The categories found in this dataset include transient chopper (T-Chop), sustained chopper (S-chop), onset (On), primary-like (Pri), phase-locked and low-frequency fibers (PLF), buildup, pauses and onset-pauser (On-Pauser).

We also examine whether certain physiological properties of the neuron, such as CF and unit type in for CN neurons affect the history parameter estimates. We classify fibers as low CF for frequencies below 1 kHz, mid-CF for frequencies between 1 and 3 kHz and high CF for frequencies above 3 kHz (Figure 29B and C). For these categories, the only significant difference between parameter estimates is seen in the 9-10 ms bin (Kruskal-Wallis test $\chi^2(2,74) = 6.71, p = 0.0349$). Posthoc multiple comparisons analysis revealed that the low-CF and high-CF groups differed at a 5% level of significance. In the CN, the three groups only differ in the 50-100 ms bin (Kruskal-Wallis test $\chi^2(2,80) = 23.7, p \ll 0.001$), with posthoc multiple comparisons analysis revealing that the high-CF group significantly differed from both the low and mid-CF groups at a 5% level of significance. This difference is likely due to the slightly excitatory activity (multiplicative effect greater than 1) seen in the 50-100 ms bin estimates for the high-CF group. Overall, the history effect is very similar across CF.

Similarly, unit type in the CN does not play a significant role in history parameter estimates. CN neurons are typically classified into unit types according to their response to short tone-bursts (Pfeiffer, 1966; Bourk, 1976; Rhode and Smith, 1986; Blackburn and Sachs, 1989). In this dataset, approximately 1/3 of the fibers were previously classified by P. Cariani

into their respective unit types, and the history parameters for each unit type are shown in Figure 29D. Kruskal-Wallis testing did not reveal differences between the 8 groups of neurons in any of the history bins at a 5% level of significance. However, this analysis suffers from a small number of classified neurons in each group. For example, only one neuron was classified as Buildup, and only two as Onset-pauser. It is possible that significant effects with unit type could be observed with a larger dataset.
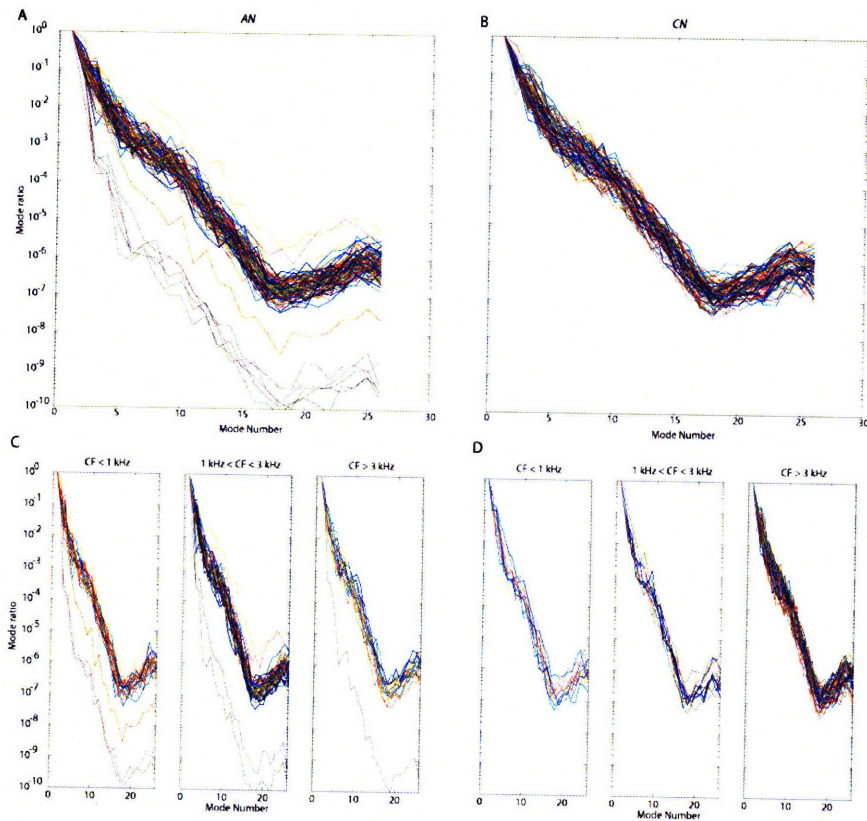

Figure 30: Ratio of magnitude of the largest parameter value for each mode number and the largest overall parameter value for all the mode numbers. A. Mode ratio for entire AN dataset. B. Mode ratio for entire CN dataset. C. Mode ratio for AN dataset with neurons categorized by their CF into three groups: those with CF less than 1 kHz, with CF between 1 and 3 kHz and with CF greater than 3 kHz. D. Same as D for CN dataset.

To summarize the parameter estimates of the EMD decomposition portion of the model, we observe the magnitude of the parameter estimates at each mode. To do so, we compute a metric that we refer to as the mode ratio, the ratio of the magnitude of the largest parameter estimate at each mode and the magnitude of the largest parameter estimate overall. The mode ratio as a function of mode number is shown for the AN and CN datasets in Figure 30 A and B, respectively. The largest parameter estimates are typically seen for the first or second modes for these datasets, and therefore, the mode ratio at mode 1 or 2 is typically 1. For modes after

mode 1 or 2, the mode ratio decreases linearly on a logarithmic scale, indicating an exponential decay in the magnitude of the parameter estimate for each subsequent mode on a linear scale, until approximately mode 17. The magnitudes of the largest parameter estimates for modes 17 through 26 are fairly constant, and even increase slightly with mode number. The shape of the mode ratio curves are similar for AN and CN, although the CN curves tend to be somewhat more homogenous. A Wilcoxon ranksum test revealed that differences between AN and CN mode ratios for each mode are only significant at the 10th ($p = 0.0376$) and 26th modes ($p = 0.009$). There, the patterns of the EM parameter estimates, as indicated by the relative size of the magnitude of the largest parameter estimates at each mode, are very similar across the AN and CN datasets for this sentence stimulus.

To examine whether the mode ratio curves in the AN and CN carry a CF dependence, we again group the cells into the low, mid and high-CF groups defined earlier in Figure 29. The non-parametric one-way analysis of variance Kruskal-Wallis test rejected the hypothesis that the samples at each mode came from distributions with equal medians ($\chi^2(2, 82) = 6.16, p = 0.0459$) only for the 20th, but not for the other modes at a 5% level of significance. Posthoc multiple comparisons analysis revealed that the low-CF and mid-CF groups have significantly different mode ratios for at the 20th mode at a 5% level of significance. Overall, there is not much evidence for CF-dependence in the mode ratios for the AN dataset. However, the CN dataset does show evidence for a CF-dependence in the mode ratios. In the CN dataset, all but the 23rd mode, the Kruskal-Wallis test rejects the hypothesis of no-CF dependence on the median values of the mode-ratio distributions at a 5% level of significance. Posthoc multiple comparisons analysis reveals that the high-CF group always differs from the low or mid-CF groups, or both, although the low and mid-CF groups do not differ from one another for any of the modes. Because certain unit types in the CN are correlated with a specific CF range (i.e. Rhode and Smith, 1986), and 2/3 of the neurons in the CN dataset are unlabeled, this analysis could be repeated for a larger set of labeled CN responses to examine whether the CF

dependence of the mode-ratio seen in this analysis depends on the unit type. It is possible that another metric besides the ratio of the magnitude of the mode parameters would yield a difference between the AN and CN populations and a more significant CF dependence in the AN dataset, but those differences may be obstructed here by the size of the largest mode parameter estimates.

4.2.3   Goodness-of-fit assessment for entire auditory nerve and cochlear nucleus dataset

To assess the goodness-of-fit of the auditory nerve dataset in the GLM-EMD framework, we use KS plots and look at the best and worst models as well as the relative KS statistics for the entire dataset. The best KS plots for the auditory nerve dataset (Figure 31A) are very close to linear and do not deviate at all from the 95% confidence interval (relative KS statistic < 1). For this dataset, a different response has the best KS plot for each of the three models; however, the responses deemed as those having the best KS plot for one model also have a relative KS statistic < 1 when another model is used (not shown). As in the GLM-gammatone example, an important similarity between the units that have the best KS plots for each of the models is that the baseline response is fairly low, and therefore the confidence bounds of the KS statistic are fairly large. On the contrary, the response having the worst KS statistic (Figure 31B) has a very vigorous response, and the resulting confidence bounds are very small (between 5 and 7 depending on the model). However, notice that the shape of the KS plot is rather linear, indicating that the models do a relatively fine job predicting the probability of each spike. Because the KS plot is a cumulative measure, the deviation from the confidence intervals for the small intervals indicates that the deviation is restricted to only a small range of intervals. For all three models, approximately 35% of the responses fall entirely within the 95% confidence intervals and the majority of the responses fall within 2 times the 95% confidence intervals (Figure 31C). Because the worst KS plot is such an isolated outlier on each histogram, while the shape of its KS plot is quite satisfactory, we conclude that the overall goodness-of-fit of the models as indicated by the KS plots is satisfactory for the entire dataset in predicting the

probability of the consequent spike on a spike by spike basis. It is not surprising that the range of the relative KS statistics is similar using the GLM-Gammatone model (Figure 25) and the GLM-EMD model, supporting the common hypothesis that the response in the auditory nerve can be well characterized by one bandpass (gammatone) filter and delays. Our findings here support this notion.
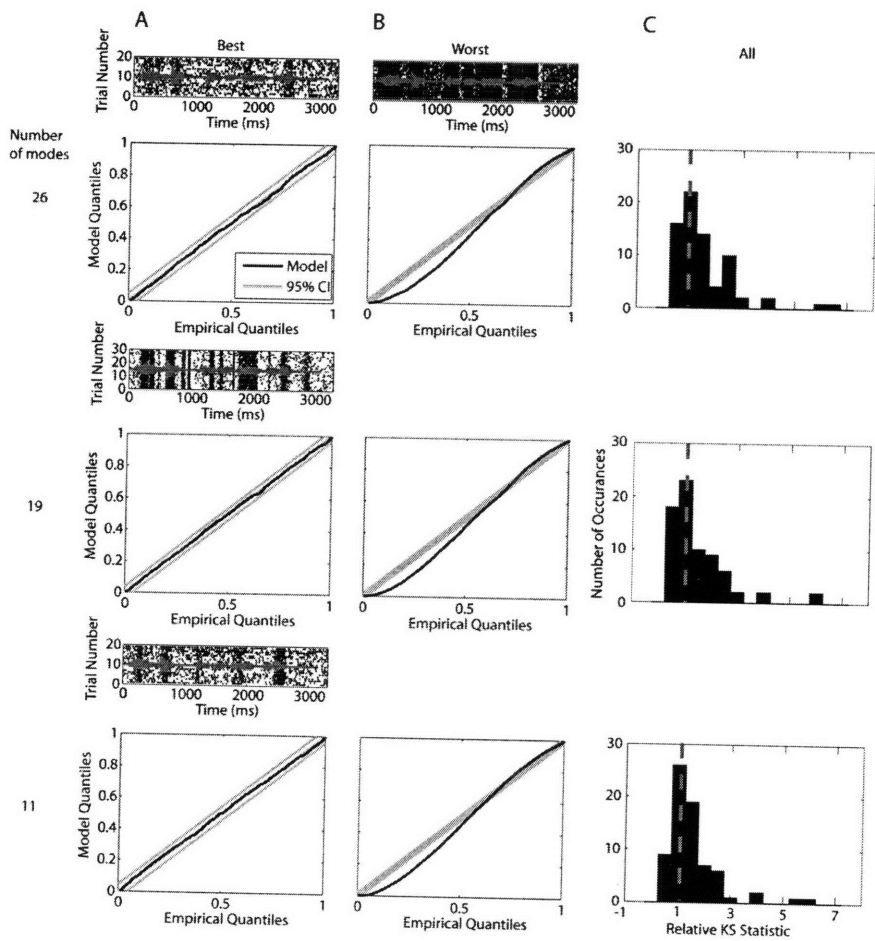

Figure 31: Range of KS fits in all auditory nerve data in the dataset. Responses and histograms for each of three models: 26 total modes (top rows), 19 total modes (middle rows) and 11 total modes (bottom rows). A. Best responses as judged by the smallest deviation from the diagonal (best KS statistic). The responses having the best KS statistic arise from a different neuron for each of the three models. B. Worst KS plot as judged by the worst KS statistic. The model of the response with the worst KS statistic arises from the same neuron (top dot raster) for each of the three models. C. Range of KS fits as shown by a histogram of the relative KS statistics for the entire AN dataset. A relative KS statistic is the ratio of the KS statistic relative to the 95% confidence interval for that neuron (i.e. a KS plot whose largest deviation just reaches the top of the 95% confidence bound in one location has a relative KS statistic of 1).

The range of the goodness-of-fit for each of the three models seen in the cochlear nucleus dataset are similar to those seen in the auditory nerve dataset. Again, the response with the best KS plot has a low firing rate, although in this case, the best response is the same for each of the three models. The best response virtually lies on the diagonal and does not deviate from the 95% confidence bounds (relative KS statistic approximately .4 for each of the

169

three models). Again, the worst KS plot arises from the unit with a very high firing rate, such that the bounds of the KS plot are very small, although the shape of the KS plot for this response is fairly linear. The worst response is a far outlier on the histogram of the KS statistics relative to the 95% confidence intervals, such that approximately 40-45% of the responses have KS plots that fall within the 95% confidence intervals for each of the three models, and the vast majority of the responses fall within 2 confidence bounds (approximately 65% for each of the three models). In addition, we observe that the responses with the high firing rates tend to have larger KS statistics, while the KS plots look fairly linear (not shown). We therefore conclude that the goodness-of-fit of the models to the cochlear nucleus data is also quite satisfactory.
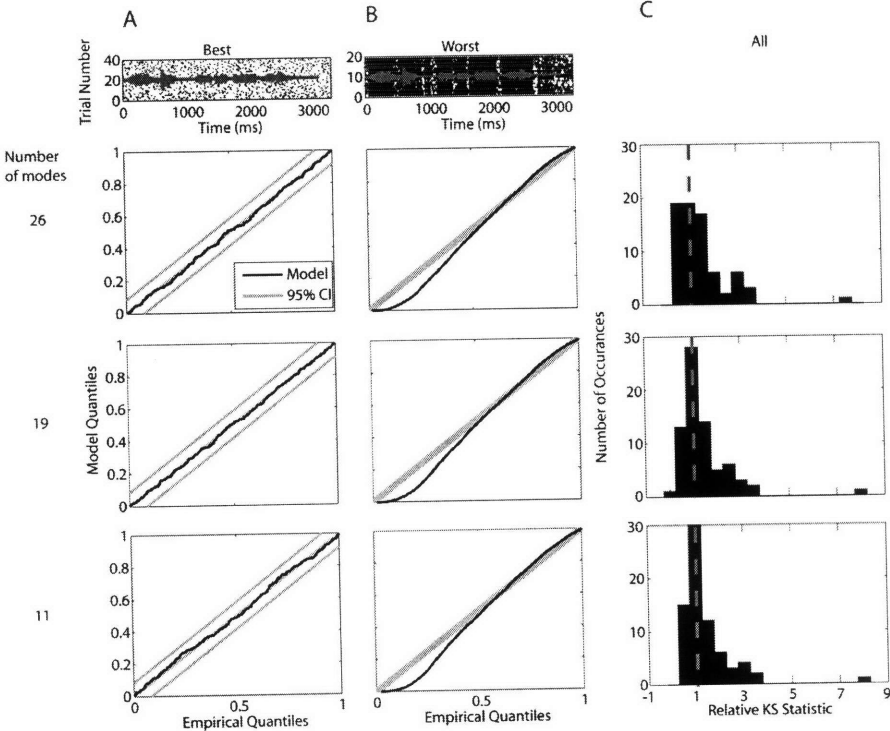

Figure 32: Range of KS fits in the cochlear nucleus. All plots same as Figure 31. However, in this case, the best KS plots in each of the three models arise from the same response (shown in top left dot raster).

4.3   Modeling of field L neurons using Empirical Mode Decomposition

The next dataset we examine contains responses from a cortical auditory area. This set of responses are a collection of responses to 20 conspecific male zebra finch songs, recorded in field L, the avian homolog of the primary auditory cortex (A1), in one isolated neuron. This dataset is interesting for two reasons. First, it allows us to examine how parameter estimates

may differ in the periphery and in the central auditory system. Second, it allows us to conduct discrimination analysis to determine whether sufficient differences in the neural responses exist to discriminate between the 20 songs with a high degree of accuracy.

4.3.1 Example of model using field L response

First, we examine an example response from the dataset. The response shown here corresponds to a conspecific song stimulus of approximately 1800 ms. The song stimulus and 10 trials of the response are shown in Figure 33A. Again, we use EMD analysis to decompose the song stimulus. For this example, we extracted 28 modes, with a ratio of lowest-order mode to highest-order mode approximately 100 to 1. Figure 33G shows the EM decomposition and the relative sizes of each of the modes. The spectra of each mode is shown in Figure 33I. As previously observed, the highest-order modes have frequency content restricted to the very low frequencies, while the lowest-order modes have broad frequency content. Since these modes are non-stationary, looking at the spectra averaged over time smears the instantaneous frequency content, although as previously stated, the instantaneous frequencies are narrowband (Huang et al., 1998). As described for the two more peripheral datasets, each of the modes are decomposed at the original sampling rate of the song stimulus (32 kHz). Next, the modes are half-wave rectified (Figure 33H) and downsampled to 1 kHz, resulting in envelopes of the original modes.

First, we examine the goodness-of-fit of our GLM-EMD model. Here, we study two versions of the model. The first is the complete model using lagged envelopes of all 28 modes and the second uses lagged envelopes of a subset of the modes from mode 1 to modes 21. We selected the second model for two reasons. First, for this example, the AIC (Figure 33J) shows a large drop at the model containing 21 modes, whereas the drop in AIC for each smaller model relative to the previous is much smaller than this drop, indicating a potential tradeoff between the number of parameter and goodness-of-fit of the model. Second, the range of number of modes for the entire stimulus set of 20 responses spans 21 to 32 (where the number

171

of modes typically increases with the song length), and we select a subset model for which the eliminated modes are small for the entire range of the data.
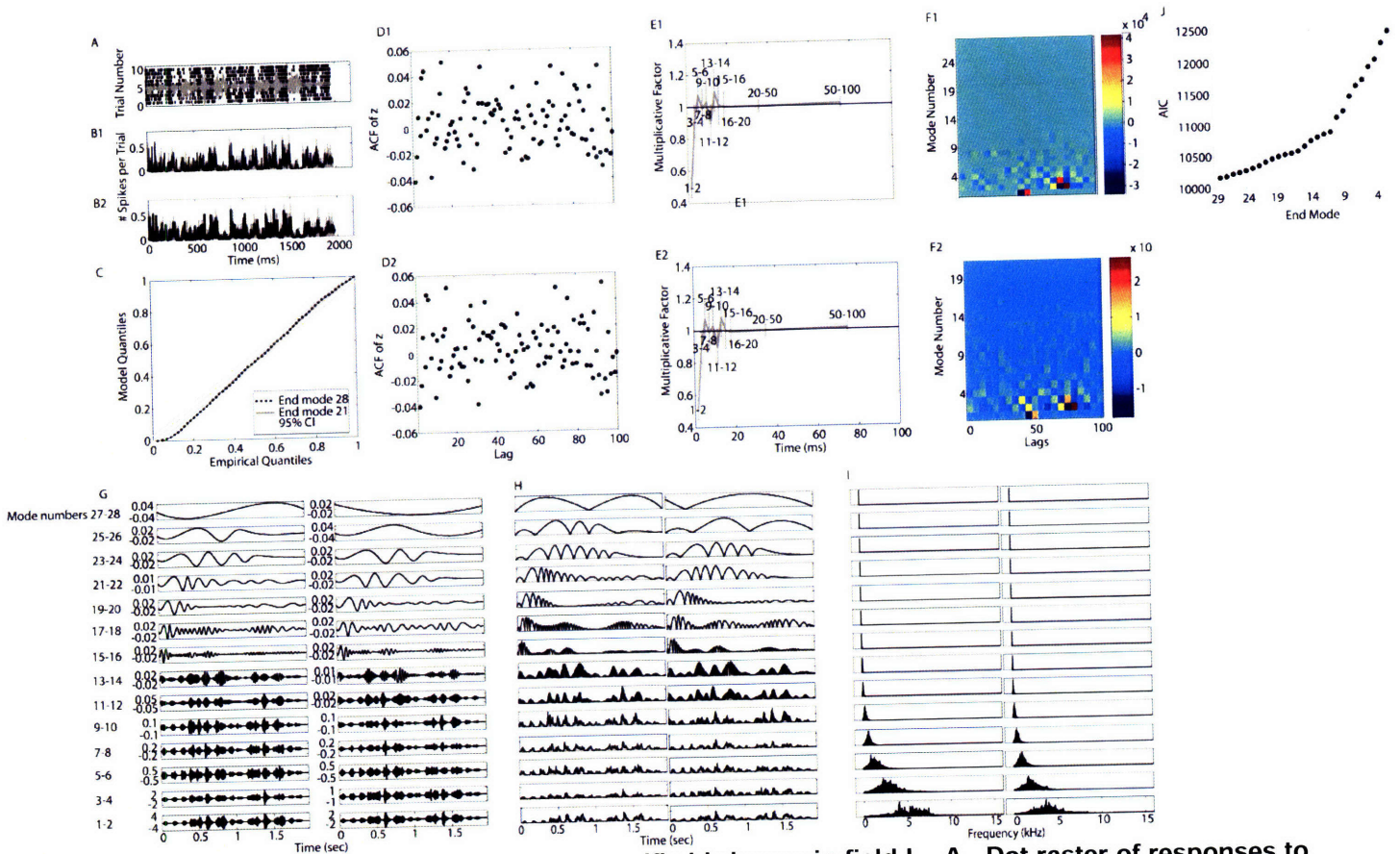

Figure 33: Example response to a conspecific bird song in field L. A. Dot raster of responses to 10 trials of bird song (shown in dark gray). B1-2. PSTH (light gray) and sum of conditional intensity function over trial (black) for response shown in A for model ending at mode 28 (B1 -- 28 modes total) and model ending at mode 21 (B2 -- 21 modes total). C. KS plot for each of the two models. The 95% confidence intervals for the KS statistic are shown in light gray. D1-2: Autocorrelation function of Gaussian rescaled intervals for each of 2 models. Light gray lines show the 95% confidence bounds. E1-2. Estimates of multiplicative history components for models using all modes (E1) and neglecting the highest-order 7 modes (E2). F1-2. Estimates of lagged parameters for model ending at mode 28 (28 models total), and at mode 21 (21 modes total). Lags used are from 0 to 100 ms using 5 ms bins. History effect parameters are used for the following history bins: 1-2, 3-4, 5-6, 7-8, 9-10, 11-12, 13-14, 15-16, 16-20, 20-50, 50-100 ms. G. IMFs for modes 1 through 28. H. Rectified IMFs for all modes for this song. I. Spectrum of IMFs for this song. J AIC for all models starting with only 2 modes (1 and 2) and with each additional mode through model with all 28 modes.

The KS plot in Figure 33C shows the uniformity of the rescaled intervals. For this example, there is some deviation from the 95% confidence intervals for the lower interspike intervals. However, because the plot is cumulative and the shape is generally linear, this

deviation is rather small and isolated to a small range of interspike intervals. The autocorrelation of the Gaussian-rescaled intervals (Figure 33D1 and D2) does not exhibit a pattern in the spread of the rescaled intervals and the number of points outside the 95% confidence intervals is less than 5% of the points calculated, indicating that this test fails to show that the rescaled intervals are correlated, and therefore fails to show that the rescaled intervals are dependent up to lag 100. We therefore conclude that the goodness-of-fit of the model is highly adequate.

The history parameter estimates for this example closely follow the pattern observed in the peripheral example shown in Figure 27 for the auditory nerve data. Here, the only significant parameter values of the history components are those that describe spiking 1-2 ms following a previous spike and result in approximately a two-fold decrease in the propensity to spike. Other parameter estimates show smaller increases and decreases, however, none of the other estimates are significant at a 95% level of significance for either of the models.

Unlike the examples of responses seen in the auditory nerve and the cochlear nucleus, the parameter estimates of the effect of the lagged envelopes of the modes tend to be less restricted to the lowest-order one or two modes. Rather, here, the parameter estimates of the largest magnitude are those that span 3-4 modes and are restricted to two time lags in the envelopes, one at 50 ms and the other at 80 ms. In general, this pattern of larger parameter estimates for a greater number of modes is seen more frequently in the field L data than in the auditory nerve or cochlear nucleus dataset.

4.3.2 Goodness-of-fit assessment for entire field L dataset

The goodness-of-fit of the field L dataset as analyzed by the KS plots is shown in Figure 34. Here, we show the best KS plot of the 20 responses in C and the corresponding response in A. The song with the best KS plot is the same for the model using all modes as well as for the smaller model excluding the 7 highest-order modes. In both cases, the KS plots are virtually indistinguishable. The KS statistics of the best KS are slightly greater than 1, indicating that the

173

plots deviate from the confidence bounds very slightly for the small interspike intervals. However, the overall shape of the plot is quite linear. The worst KS plot of the 20 responses in the dataset is shown in Figure 34D and the response is shown in Figure 34B. Again, the same response produces the worst KS plot for each of the two models. Although the KS statistic is about 3, the KS plot is quite linear, indicating that the deviation from the model is in a small range of low interspike intervals.
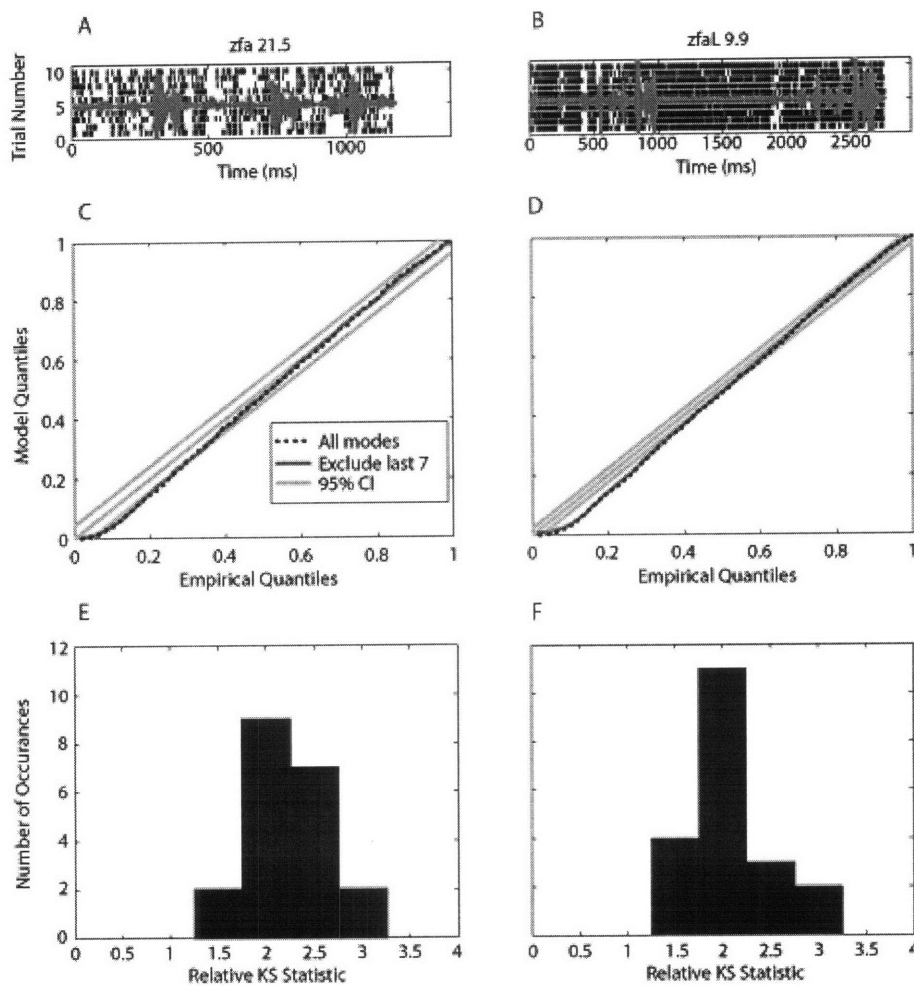

Figure 34: Range of KS fits in field L for all 20 songs in dataset. Responses (A & B) and KS plots (C & D) for best (C) and worst (D) KS plot in the dataset for each of the two models (using all modes and excluding the last 7 highest-order modes). C & D. Range of KS fits as shown by a histogram of the relative KS statistics for the entire field L dataset for the model using all modes (E) and the model neglecting the 7 highest-order modes (F). A relative KS statistic is the ratio of the KS statistic relative to the 95% confidence interval for that neuron (i.e. a KS plot whose largest deviation just reaches the top of the 95% confidence bound in one location has a relative KS statistic of 1).

As observed in the AN and CN datasets, the responses with the worst KS plots are those with very high firing rates and therefore smaller confidence bounds, while those with the best KS plots have more moderate firing rates and larger confidence bounds. Observing the histograms of the relative KS statistics for each of the two models in Figure 34E and F, we

notice that the range of relative KS statistics is between 1 and 3. Combined with the linearity of the worst KS plot, we find the fits of each of the GLM-EMD models satisfactory for the field L dataset.
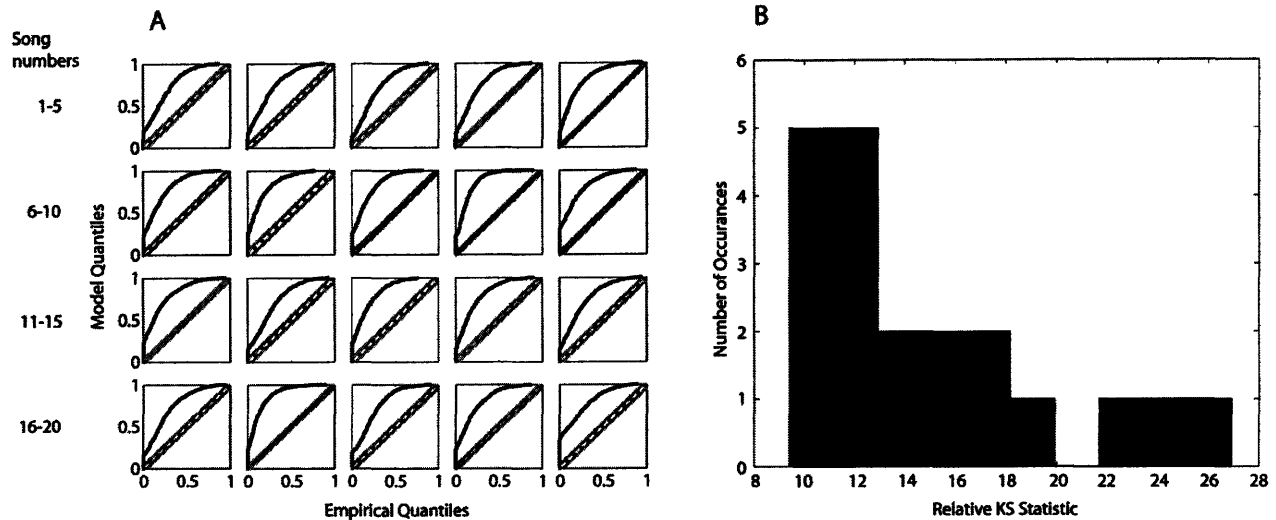

Figure 35: Goodness-of-fit of estimated responses using the spatio-temporal receptive field model as given by KS statistics. A: KS plots for the estimated responses for the 20 songs in the field L dataset using the Theunissen STRF algorithm (Theunissen et al., 2000). Briefly, the stimulus is decomposed using Gaussian filters with 250 Hz widths (standard deviation) with center frequencies between 250 and 8,000 Hz. The log of the amplitude envelope of the signal in each frequency band is taken. The STRF for each filter i, is $h_i(t)$ such that $r_{pre}(t) = \sum_{i=1}^{nf} \int h_i(\tau)s_i(t - \tau)d\tau$ and $h_i(t)$ is selected such that the predicted firing rate, $r_{pre}(t)$, is as close as possible, in the least square sense, to the peristimulus time histogram , $r(t)$. See Theunissen et al., 2000 for details on the L1 regularization involved in this calculation. B: KS statistics derived from the plots shown in A.

One popular methodology of modeling auditory responses is the spatio-temporal receptive field (STRF) estimates. These STRF predict an estimate of the PSTH of the response using shifted envelopes of the bandpass filtered sound stimulus using linear filtering theory with L1 regularization. It is possible to evaluate the goodness-of-fit of these estimates using KS plots and KS statistics by assuming that the estimate of the firing propensity at every trial is the same, as estimated by the PSTH. Because the STRF estimate of the PSTH does not use spiking history, the assumption is that the firing rate of the trial-to-trial responses should not differ. Figure 35A shows the KS plots for the estimated firing propensities of the 20 units in the field L dataset using the STRF model. Unlike the KS plots using the GLM-EMD framework, the plots are not linear, but in fact bowed, indicating that the model is not able to satisfactorily predict the

175

firing propensity of the response on a spike-by-spike basis. When the relative KS statistics are computed, the poor goodness-of-fit of the STRF framework is made even more apparent. The range of relative KS statistics, from 10 to 26 is much larger than the range seen using the GLM-EMD model in field L (approximately 1 to 3), mirroring the observation that the KS plots indicate a poor model fit to the data.

4.4.   Decoding of responses in field L based on GLM-EMD model

One motivation for creating and estimating well-fitting models that predict the spiking propensity of the cell is to gauge to what extent the stimulus can be decoded from the response, based on the estimated model. To date, this issue has been explored in the field L, using these and other similar responses to conspecific bird songs by Kamal Sen and colleagues with models involving heavyside weighted functions (Narayan et al., 2006; Wang et al., 2007). In this analysis, we wanted to explore whether incorporating the rate-based and temporal dynamics of the response inherent in the GLM-EMD model would produce a high level of discrimination between songs while treating the neural responses as point processes.

To conduct the decoding analysis, we estimate the GLM-EMD model for the 20 songs based on 9 of 10 trials and left the last trial that is not included in model fitting for the validation dataset (see Methods). Therefore, for each of 10 trials for each song, 10 GLM models are fit to the data based on 9 trials. Next, for each response in the validation dataset, we calculate the likelihood that the trial originated from each of the 20 models for all the songs. Examples of these likelihoods are shown in Figure 36. Figure 36A shows the likelihood of a trial from song 1 originating from each of the other 20 songs and Figure 36B shows the likelihood for a trial from song 8 originating from models of songs 1 through 20. Both likelihoods are extremely sharp, with one peak only, and the maximum of the likelihood centered on the correct song model. The sharpness of the likelihoods is characteristic of this analysis and is more clearly shown for 1 trial originating from each of the 20 songs in Figure 36C. Here, the most likely models are

176

correctly attributed to each song trial in all but 3 songs, which corresponds to a detection accuracy of 85% for these 20 trials (1 from each of 20 songs). With the exception of several infrequent and isolated incidents, the likelihood functions are very steep and unimodal, as those shown here.
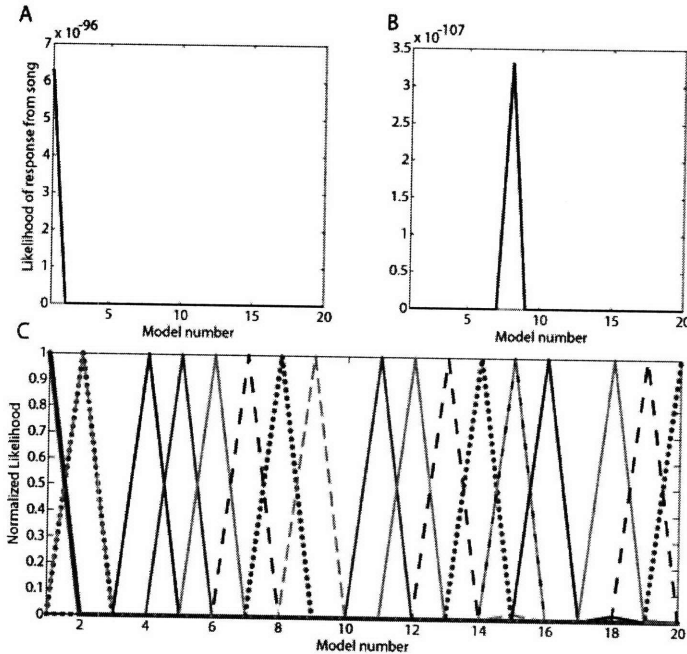

Figure 36: Example of sharpness of likelihood curves generated by decoding analysis. A. Likelihood of trial 1 of song 1 originating from song described by models 1 through 20. B. Likelihood of trial 1 of song 8 originating from song described by models 1 through 20. C. Normalized likelihood functions for trial 1 from songs 1 through 20 originating from songs described by models 1 through 20. Likelihoods are normalized between the maximum and the minimum for each model. All trials, except those from song 3, 10 and 17, are correctly discriminated.

The likelihood functions shown in Figure 36 are computed for one trial only and as a function of one song/model length. We show confusion matrices to view all detection trials for all 20 song models and various song lengths in Figure 37. For example, Figure 37A1 shows the GLM-EMD model discrimination for a full model (all modes included) for each of the 20 songs and each of the 10 trials for each song. The diagonal elements correspond to trials that originated from, and are detected as originating from, the correct song. The darkness indicates the number of trials assigned to each song (of the 10 trials for each song in the set). The song length in this example corresponds to 826 ms, the length of the shortest song in the dataset. For songs/models longer than 826 ms, only the first 826 ms are included in the analysis to obtain the confusion matrix shown in Figure 37A1. The confusion matrix indicates that some trials for songs 2, 3, 12, 16 and 18 are incorrectly identified with the model for song 13, the trials from 17 with the model for song 11 and the trials from song 8 with the model for song 1. All

177

other trials are correctly identified with the appropriate song model. For this matrix, the proportion of correctly discriminated songs is $\frac{12.1}{20}$ corresponding to 60.5%. A random guess would lead to a discrimination value of 5%, since a random guess would be as likely to assign a model to any one of the 20 songs. Therefore this discrimination analysis indicates that for a song length of 826 ms in this dataset, the complete model can correctly assign 60.5% of the trials to their correct song origin. We repeated this analysis by taking successively longer song lengths by excluding the shortest songs. The proportion of corrected songs increased to 79.5% for a song length of 1178 ms (Figure 37A2), to 82.4% for a song length of 1226 ms (Figure 37A3), and to 92.9% for a song length of 1416 ms (Figure 37A4).
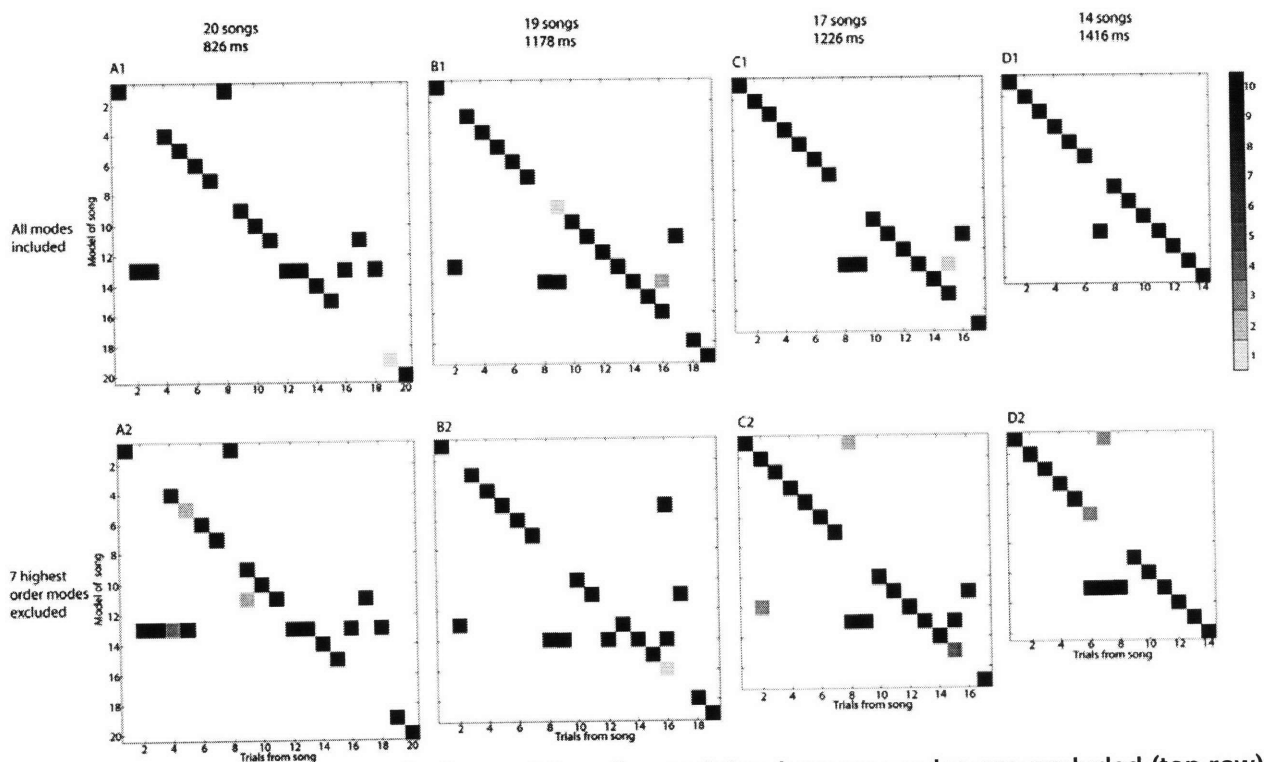

Figure 37: Comparison of confusion matrices for models where no modes are excluded (top row) and the highest-order 7 modes are excluded (bottom row) for songs of various lengths. The confusion matrices represent the most likely response trial (x-axis of each confusion matrix) to correspond to the model of each song (y-axis of each confusion matrix) for each of the 10 trials in the response. To compute these matrices, for each song, each trial is systematically excluded, and a model is fit to the other 9 trials, such that each of the 10 trials is excluded once with a corresponding model based on the other 9 trials (for a total of 200 models for all 20 songs). Next, given the 10 trials excluded from each of the 20 song models, the most likely trial to arise from each of the 20 song models is selected and marked on the confusion matrix. Elements on the diagonal represent correct discrimination, while elements off the diagonal represent incorrect discrimination. Because each of the 20 songs (and consequent responses) differed in length, the discrimination is computed for the first M ms of each song, where M represents the length of the

178

shortest song. The leftmost confusion matrices (A1 and A2) represent the discrimination performance for all 20 songs using the first 826 ms of the song. The next confusion matrices on the right (B1 & B2) are computed with the shortest song removed (19 songs total) and with a length of 1178 ms, corresponding to the length of the next shortest song. Likewise, each of the next shortest songs are excluded, such that the length over which the decoding is computed is 1226 ms (C1 & C2) and 1416 ms (D1 & D2).

We repeated the analysis for partial models for which lagged envelopes of the last 7 modes are excluded. The discrimination typically degraded somewhat from that using the full models, with more instances of confusion, evidenced as an increase in non-diagonal elements in the confusion matrices relative to those calculated with the full models. For example, for the shortest song length of 826 ms, trials from songs 4 and 5 are more often confused with the model from song 13 and trials from song 9 with the model for song 11. These confusions did not occur with the full model. The rates of detection for the four example song lengths show in Figure 37 are slightly reduced to 60.5%, 68.4%, 78.2% and 80% for song lengths of 826, 1178, 1226 and 1416 ms, respectively, for the reduced models (Figure 37B1, B2, B3 and B4). However, the increase with song length is nevertheless evident.
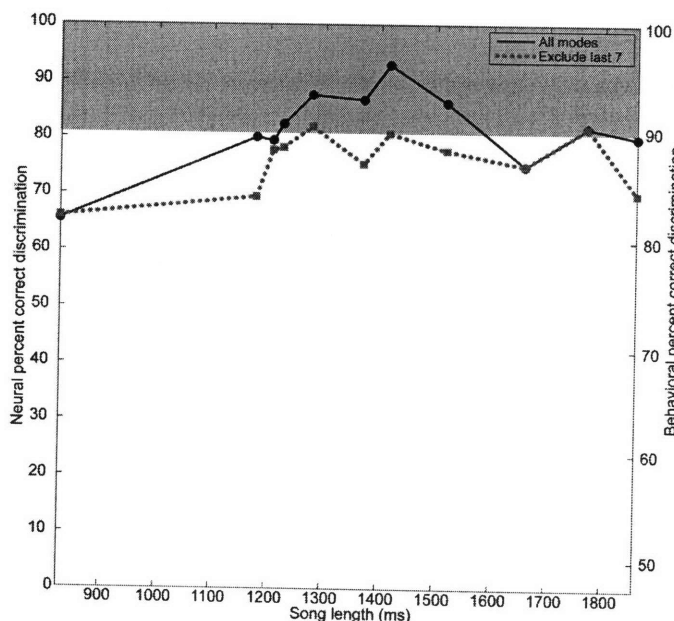
Figure 38: Comparison of neural decoding performance within the likelihood based framework and behavioral performance of zebra finches discriminating conspecific songs (Cynx, 1993; Shinn-Cunningham et al., 2006). Adapted from Wang et al. (2007). Discrimination performance is shown (based on results from confusion matrices shown in Figure 37) as a function of the song length for all modes (black solid) and the last 7 highest-order modes excluded (gray dotted). Behavioral performance shown is average performance for all song lengths.

A summary of the proportion of correctly detected trials for the full and reduced models are shown in Figure 38A using the GLM-EMD framework. The results from the 8 confusion matrices shown in Figure 37 are included as data points on the plot. To construct the plot, we

179

varied song length from 826 ms (all 20 songs), to 1863 ms (10 songs). The proportion of correctly detected trials varies from 60.5% to 92.9% for the full model and from 60.5% to 80% for the reduced model. As seen in the patterns with the confusion matrices, discrimination analysis is at least as good with the full model than the reduced model, and is typically better.

We compared our results with the behavioral performance of birdsongs discriminating between conspecific songs (Cynx, 1993; Shinn-Cunningham et a., 2006). These studies found discrimination in the range of 90-100%. However, in these studies, the chance level for discrimination was 50%, while our study, chance level is 5%. We therefore used a method common to psychoacoustics to plot the neural and behavioral performance on the same axis (Wang et al., 2007; Shinn-Cunningham et al., 2006), with the behavioral performance shown in gray in (Figure 38). Comparing the discrimination performance using out methodology to the behavioral performance, Figure 38 shows that the neural performance is within the range of the behavioral performance for this neuron. These results indicate that the responses have sufficient information captured by our decoding strategy to discriminate between the different songs at the level of one neuron (see Discussion).

5. DISCUSSION

5.1 Summary

Neural spiking is a dynamic stochastic process that is mediated by extrinsic covariates, the neuron's own past spiking history, and potentially other factors, such as the spiking of nearby neurons. Although neural spiking has been explored as a stochastic point process for many sensory systems (visual Brown et al., 2002; somatosensory Haslinger et al., 2007; motor Brillinger, 1988, Eden et al., 2004, Paninski, 2004, Truccolo et al., 2005), point process models of auditory responses including effects of spiking history and stimulus-related covariates are not common. In this work, we presented a methodology and modeling examples of responses in three auditory centers to natural stimuli: auditory nerve and cochlear nucleus in response to a sentence and field L in response to conspecific bird songs. To model these natural stimuli, we used a recent decomposition technique of EMD suitable for creating a basis function of nonlinear, non-stationary signals. We used time-shifted versions of these basis functions, representing the stimulus subject to delays, and the neuron's past spiking history to explore the effect of these covariates on the neural spiking in a likelihood-based GLM framework. In interpreting the responses in the three nuclei, we illustrated that the responses in the two more peripheral nuclei are similar, both in terms of the estimates of the effect of the stimulus and the spiking history, while those in field L differ somewhat with respect to the stimulus effect. We also found that the neurons in the auditory nerve and cochlear nucleus exhibit refractory behavior for spiking history up to 14 ms in the past. Highly adequate goodness-of-fit was illustrated for our models using criteria appropriate for point process data, such as KS plots, autocorrelation analysis and AIC. Lastly, using the birdsong responses, we performed decoding analysis to measure the level of discrimination performance capability of the neuron under our modeling scheme. We found that the neural discrimination using our model that treats the data as a point process and uses likelihood methods for decoding is within the range of behavioral discrimination performance observed for this species.

5.2. Current point process framework presents important extensions and advantages to previous auditory point process models

The earliest point process models, proposed by Siebert, Gray and Colburn (Siebert and Gray, 1963; Siebert, 1965; Gray, 1967; Colburn, 1969; Colburn, 1973), acknowledged the stochastic nature of firing patterns in the auditory nerve and characterized the response as a non-homogeneous Poisson process. Gaumond et al. (1982) suggested that homogeneous Poisson firing patterns are not empirically consistent with the firing patterns seen in the auditory nerve, but that a conditional intensity function capturing both the peristimulus time histogram and the previous spiking history as a multiplicative model was more consistent with the empirical data. Using this multiplicative model of the CIF, Johnson and Swami (1983) and Miller and Mark (1992) showed empirical model results of the PSTH corrected for the effects of the previous spiking history in the AN, and Johnson and colleagues showed analogous findings in the lateral superior olive (Johnson et al., 1986; Turcott et al., 1994; Johnson et al., 1996). A commonly-used model of the auditory periphery (Heinz, 2001a, 2001b) uses earlier models (Colburn 1969, 1973; Carney, 1993) of the auditory nerve responses as a non-homogeneous Poisson process with additional nonlinearities.

These earlier point process models of auditory neurons share several similarities, but also exhibit important methodological and philosophical differences, from those we propose here. Although we also employ the multiplicative form of the conditional intensity function, our model description is much more general than those proposed by Miller, Mark, Johnson and colleagues. Whereas the stimulus response portion of their models only captures the PSTH corrected for the stimulus spiking history, our stimulus response portion can capture the effect of any arbitrary covariates related to the stimulus. For example, in the models we show in this chapter, the stimulus response portion is a either a gammatone model of the auditory nerve/inner hair cell synapse or a bank of filters representing the rectified envelopes of the EMD of the stimulus. Second, while Miller and Mark (1992) only consider the absolute and relative

refractory effects of the neuron in their recovery function, we allow for the history component to potentially longer intrinsic biophysical properties of the cell such as rebound excitation that may be seen in more central neurons. The methodologies for parameter estimation also differ between those proposed here and earlier work. Mark and Miller (1992) use the EM algorithm for parameter estimation, with the constraint of a necessarily periodic stimulus (Miller, 1985), while Johnson and colleagues empirically maximize the likelihood for their models. In our framework, we build on earlier studies of point process models in the likelihood framework (Brillinger, 1988; McCullagh and Nelder, 1989; Kass and Ventura, 2001; Truccolo et al. 2005; Czanner et al., 2008) and define the equivalent problem in the GLM framework, allowing for efficient ML solutions via iteratively reweighted least squares (McCullagh and Nelder, 1989). Moreover, the ML solution to a GLM has been implemented in MATLAB (The Mathworks, Natick, MA). Another important methodological difference is our use of the time-rescaling theorem (Brown et al., 2002) to construct goodness-of-fit assessments of our models appropriate for point process data and utilize AIC for model order selection (see section 5.4).

5.3   EMD as a new tool for decomposition of natural stimuli

Many signal processing alternatives to EMD are available in the analysis of time-series decompositions, including wavelet and Fourier methods. The direct extraction of timescales from the local characteristics of the signal is a uniqueness of the EMD methodology. In other words, the basis functions are nonlinear and non-stationary, and are directly extracted from the signal and the timescales present in the IMFs are dictated by the signal, rather than pre-determined. An alternative strategy for analyzing non-stationary time series are short window Fourier methods. However, the basic functions for Fourier analysis are restricted to windowed sinusoidal functions and necessitate pre-determined cutoff frequencies for separating the signal into different frequency bands. However, this restriction can be circumvented, as it is in the cochlea, with overlapping filters.

Although wavelet analysis is not restricted to sinusoidal functions, the mother wavelet,

which determines the scaled and shifted signals used in the decomposition, is pre-determined. The range of mother wavelets is large; however, no guidelines exist facilitating an appropriate choice for the wavelet function (Quatieri, 2001). The analysis could be degraded if the timescales of an inappropriately chosen mother wavelet do not match those of the signal. Further, wavelet analysis uses linear, time-invariant techniques, while IMFs are nonlinear and non-stationary. Therefore, wavelet analysis cannot be used to adapt to local changes of the signal (Huang et al. 1998).

A further advantage of the EMD methods is the sparseness of the basic functions. As illustrated in Figure 22, the two distinct timescales contained in the signal are extracted by the EMD analysis in its first two basic functions. From this standpoint, EMD analysis is more flexible than Fourier and wavelet analysis. For instance, EMD decomposition would return the same modes as Fourier decomposition of a stationary signal composed of two sine waves. The advantage of EMD over Fourier decomposition is that it resolves time-frequency information that cannot be resolved with Fourier. The advantage over wavelet decomposition is that with EMD, one is not bound by the choice of a mother wavelet. EMD thus creates a sparse representation and extracts the underlying timescales found in the signal.

One potential disadvantage of the EMD analysis as applied to the analysis of natural stimuli is directly connected to the advantage of empirically extracting timescales unique to the signal. Since basis functions are not pre-determined in terms of their cutoff frequencies, for example, the decomposition of each natural stimulus will differ. Therefore, the size of the estimated parameters representing the effect of the modes on the CIF for different birdsongs is not directly comparable. In addition, the decomposition for one stimulus cannot be used to predict the response to another stimulus. However, we suggest that for stimuli that have a similar origin, such as a collection of conspecific bird songs or a collection of sentences spoken by a single speaker, the relative range of timescales in each IMF will be similar. For instance, we showed that high numbered EMFs contain very low frequency components, while low

184

numbered IMFs contain a higher range of frequencies and more temporal variation with frequency (see Figure 26, Figure 33). This pattern holds across the stimuli we analyzed in our dataset. Therefore, one could infer coefficients related to the IMFs as related to a range of time-scales to compare to analysis for other similar stimuli.

5.4 Confirming model adequacy through goodness-of-fit criteria and model order selection metrics

In this work, we assessed model goodness-of-fit and model selection criteria using four metrics: Kolmogorov-Smirnov plots, autocorrelation analysis, AIC and decoding analysis. Since no one metric provides assessment of the appropriateness of all aspects of the model to the data, it is important to employ multiple measures. In addition, none of the analyses yield an unequivocal result regarding the appropriateness of the model. Rather, they provide the degree to which the model appropriately captures the stochastic nature of the data and help to identify what covariates may be missing from the model. For this reason as well, multiple goodness-of-fit metrics should be examined together to tradeoff between different model formulations in light of the assessments.

Kolmogorov-Smirnov plots and autocorrelation analysis are two critical assessments we employed for all the model examples. The KS plots assesses to what extent the stochastic nature of the modeled CIF accurately reflects the stochastic nature of the true neural data. Comparisons of KS plots of our models and those based on the STRF model highlighted two important differences in our framework that better capture the stochastic nature of the data. First, the STRF framework excludes history dependence by averaging the response over multiple trials. Including history dependence has been shown to increasingly capture the model variability (Truccolo et al., 2005; Czanner et al., 2008). Second, the STRF model includes regularization and smoothing parameters such that the estimated PSTH based on the STRF typically differs from the true PSTH. Both factors contribute to the degraded KS plots relative to those based on our model analysis. Additionally, autocorrelation analysis is useful in assessing

whether additional correlation exists in the residual or whether the rescaled intervals are correlated with extrinsic covariates that may have been left out of the model. We did not see evidence of autocorrelation of the rescaled intervals in our model parameters. Adequate KS and autocorrelation plots indicate that rescaled intervals are uniform and independently distributed, respectively, and give further evidence to the appropriateness of the models.

In addition to performing rigorous goodness-of-fit assessment using Kolmogorov-Smirnov plots and autocorrelation analysis, we also performed decoding analysis. Unlike the two other goodness-of-fit metrics, decoding analysis assesses how well the effect of the extrinsic covariate is captured in the model. We found that song discrimination ranges from 60% to over 90% for a set of 20 songs, depending on the length of the song stimulus, and that the extended model using all estimates modes yields somewhat higher decoding performance for a given song length than a model that excludes the last 7 estimated modes. However, previous work finds an asymptotic equivalence between AIC and cross validation analysis (Stone, 1977), although AIC generally outperforms cross validation for prediction performance (Efron, 2004). Since the larger model yielded a lower value of AIC, it is not surprising that decoding results using the larger model are superior to those using a smaller model. More generally, the high discrimination performance of the decoding analysis was also partially reflected in the KS plots, showing that all model using song stimuli closely captured the stochastic nature of the neural spike trains, and the autocorrelation analysis, illustrating that the rescaled spike time structure did not reveal correlations or the absence of some extrinsic covariate.

In addition to predicting some decoding results, AIC provides additional benefits by rewarding models for which the model likelihood closer resembled the true stochastic process generating the data and penalizing for larger models. Although in this work, we used AIC to compare between models with different numbers of modes, AIC could have also been used to examine either all history covariates added to model goodness-of-fit, or to assess whether finer

186

or coarser time-shifts of the modes could have provided improvements, as well as to assess any additional covariates represented in the model. Generally, the AIC is also a useful tool to prevent overfitting of the data by examining whether smaller models would greatly degrade from the goodness-of-fit. However, for our examples, the high ratio of model parameters to the data size (approximately 1 to 100 for the lowest-order models in the dataset), make overfitting unlikely.

Some current models of auditory data summarize the neural response by computing the PSTH, the average of the response over stimulus trials. By computing a PSTH, the experimenter makes the statement that spiking history does not significantly alter the neural response. However, it has been described that spiking history effects that are constant across trial can yield different instantiations of the neural data (Czanner et al., 2008). In addition, inclusion of spiking history improves the degree to which the stochastic nature of the response can be captured (Truccolo et al., 2005; Czanner et al, 2007). In earlier work at the level of the auditory nerve, Mark and Miller (1992) showed that excluding the stimulus history leads to PSTH variance estimates 3 times lower than those empirically observed in the data. Not including model covariates or not analyzing the effects of the covariates simultaneously may lead to an inaccurate estimate of the covariate effect. As earlier studies show, neglecting spike history information can interfere with model estimates of the stimulus effect in STA and reverse correlation analysis (Aguera y Arcas and Fairhall 2003; Aguera y Arcas et al., 2003; Truccolo et al., 2005) and the PSTH (Gaumond et al, 1982; Miller and Mark, 1992). Therefore, we suggest that more general point process models, capturing the stimulus and history-related properties of the neuron be used. For these models, goodness-of-fit metrics, such as the Pearson correlation between the true and estimated PSTH, do not fully reflect the extent to which the model captures the data variability, and certainly neglects important stochastic properties of the data.

5.5. Extending GLM models for arbitrary stimulus decomposition and phase locking to fine temporal structure of the stimulus

Our motivation in this chapter was to present a general methodology for describing auditory responses to natural stimuli. In this chapter, we presented two examples of decomposition choices of the stimulus in the GLM framework. We presented models where the stimulus decomposition component either modeled the peripheral filtering of the ear using a gammatone filter or an EMD of the acoustic stimulus. The decomposition of the stimulus can be quite general. In fact, current models using Fourier decompositions of the stimulus (Aertsen et al, 1981; Theunissen et al., 2000), Gabor functions (Qiu et al., 2003) or spherical approximation functions (Jenison et al., 1998) to represent the regularized spatio-temporal response dynamics can certainly be adapted to the GLM framework to estimate both the effects of the stimulus dynamics and the previous spiking history on the neural response.

Second, in the current implementation, we modeled the effect of the envelopes of the stimulus covariates on the neural response. We were therefore unconcerned with response related to the fine temporal dynamics of the stimulus. Some neurons in the auditory system exhibit a phenomenon called phase locking, where spike discharges do not occur at every cycle of the stimulus waveform (because of refractory effects), but at a particular phase of the stimulus cycle (Javel et al., 1968 for pure tones). In auditory nerve neurons, phase locking has been observed for sinusoids up to 5 kHz (Johnson, 1980). We could capture this effect in the stimulus-effect portion of the response by shifting the stimulus by multiples of the stimulus period (Gaumond et al., 1982). Nevertheless, it is important to point out that for cells with a refractory period of 1 ms, we would not expect to observe more than one spike in less than 1 ms. To capture phase locking in the fine structure, we would select a sufficiently small binwidth in the tens to hundreds of microsecond range, depending on the stimulus. Next, we would select stimulus shifts of small durations, sufficient to capture spikes that may equal the refractory period plus multiples of the stimulus period. However, this methodology would create

188

a large number of estimated parameters, many of which may not be significant to the model. We could then use AIC to eliminate unnecessary time bins. A look at the spectrum of the stimulus-effect could estimate the frequency of phase locking. As an alternative to creating many parameters describing the stimulus shifts, a nonparametric form of the model could be employed, similar to that suggested in Frank et al., 2002. In his work, Frank and colleagues describe the history component as a summation of splines, with variable control points to capture the effect of the previous spiking history, although here we would extend the framework to the stimulus-effect component. Again, AIC could be used to reduce the number of spline control points.

5.6 *Good decoding performance independent of filtering*

Neural discrimination performance using our ML decoding strategies is similar to behavioral discrimination performance for the same species (Cynx, 1993; Shinn-Cunningham et al., 2006). Neural discrimination performance was also measured by Wang et al. (2007) for conspecific birdsongs using Victor and Purpura (1997) and van-Rossum (2001) metrics. All the metrics used by Wang et al. involve augmentation of the original spike train. The van Rossum spike distance metric convolve the spike train with a heavyweight function and scale it by an exponential, inducing some temporal smearing and transforming the discrete point process into a continuous valued signal. The Victor and Purpura metrics instead add, remove and change the interval between spikes to calculate a cost function related to a distance between the spikes, with the metrics of the cost function related to the methods used to augment the spikes, although one could imagine employing other cost functions. Wang et al. conclude that the van-Rossum (2001) metric produces discrimination performance in their best neuron on par with behavioral levels, and that some form of temporal smearing with a timescale of approximately 10 ms is likely to be used in the birdsong discrimination circuitry.

Our methodology also produces discrimination performance at the level seen behaviorally. However, our methodology does not augment spike trains as does the spike

distance metrics, and rather uses methods appropriate for point process data to calculate the most likely song model that could have generated the specific spike train. The performance of our methods suggests that temporal smearing is not necessary to obtain accurate discrimination performance. Further, the optimal time scale of 10 ms resolution suggested by the Narayan et al. (2006) and Wang et al. (2007) work may not be a behaviorally necessary property of the neural network that underlies performance, but rather a consequence of the analysis methods used.

ACKNOWLEDGEMENTS

6.   REFERENCES

Aertsen AM, Johannesma PI (1981) The spectro-temporal receptive field.   A functional characteristic of auditory neurons. *Biol Cybern* 42:133-143.

Aguera y Arcas B, Fiarhall AL (2003) What causes a neuron to spike? *Neural Comput* 15: 1789-1807.

Aguera y Arcas B, FairhallAL, Bialek W (2003) Computation in a single neuron: Hodgkin and Huxley revisited. *Neural Comput* 15: 1715-1749.

Akaike H (1973) Information theory as an extension of the maximum likelihood principle. In: *Second International Symposium on Information Theory*, edited by Petrov BN and Csaki F. Budapest: Akademiai Kiado, p. 267-281.

Anderson DJ, Rose JE, Hind JE, Brugge JF (1971) Temporal position of discharges in single auditory nerve within the cycle of a sine-wave stimulus: frequency and intensity effects. *J Acoust Soc Am* 49: 1131-1139.

Bendor D, Wang X (2005) The neuronal representation of pitch in primate auditory cortex. *Nature* 436: 1161-1165.

Blackburn CC, Sachs MB (1989) Classification of unit types in the anterovental cochlear nucleus: PST histograms and regularity analysis. *J Neurophysiol* 62: 1303-1329.

Boer de R, Kuyper P (1968) Triggered correlation. *IEEE Trans Biomed Eng* 15: 169-79.

Bourk TR (1976) Electrical responses of neural units in the anteroventral cochlear nucleus of the cat. Doctoral Dissertation. Dept. of Electrical Engineering and Computer Science, MIT, Cambridge, MA.

Box GEP, Jenkins GM, Reinsel GC (1994) *Time series analysis, forecasting and control.* 3rd ed. Englewood Cliffs, NJ: Prentice-Hall.

Brillinger DR (1988) Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol Cyber* 59: 189-200.

Brown EN (2005) *Theory of Point Processes for Neural Systems.* In: Chow CC, Gutkin B, Hansel D, Meunier C, Dalibard J, eds. Methods and Models in Neurophysics. Paris, Elsevier, Chapter 14: 691-726.

Brown EN, Barbieri R, Eden UT, Frank LM (2003) *Likelihood methods for neural data analysis.* In: Feng J, ed. Computational Neuroscience: A Comprehensive Approach. London: CRC, Chapter 9: 253-286.

Brown EN, Barbieri R, Ventura V, Kass RE, Frank LM (2002) Time-Rescaling theorem and its application to neural spike train data analysis. *Neural Comput* 14:325-346.

Brown EN, Frank LM, Tang D, Quirk MC, Wilson MA (1998) A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J Neurosci* 18: 7411-7425.

Brown EN, Nguyen DP, Frank LM, Wilson MA, Solo V (2001). Analysis of neural receptive field plasticity by point process adaptive filtering. *Proc Natl Acad Sci* 98: 12261-66.

Cant NB (1992) The cochlear nucleus: neuronal types and their synaptic organization. In *The mammalian auditory pathway: neuroanatomy.* Webster DB, Popper AN, Fay RR (Eds). New York: Springer-Verlag: 66-116.

Carney LH (1993) A model for the responses of low-frequency auditory-nerve fibers in cat. *J Acoust Soc of Am* 93: 401-417.

Carr CE, Konishi M (1990) A circuit for detection of interaural time differences in the brain stem of the barn owl. *J Neurosci* 10: 3227-3246.

Cariani P and Delgutte B. Neural correlates of the pitch of complex tones. I. pitch and pitch salience. J. Neurophysiol. 1996 76:1698-1716.

Colburn HS (1969) Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Colburn HS (1973) Theory of binaural interaction based on auditory nerve data. I. General strategy and preliminary results on interaural discrimination. *J Acoust Soc Am*, 54: 1458-1470.

Cooper NP, Robertson D, Yates GK (1993) Cochlear nerve fiber response s to amplitude-modulated stimuli: variations with spontaneous rate and other response characteristics. *J Neurophysiol* 70: 370-386.

Cynx J (1993) Conspecific song perception in zebra finches (*Taeniopygia guttata*). *J Comp Psychol* 107: 395-402.

Czanner G, Dreyer AA, Eden UT, Wirth S, Lim HH, Suzuki WA, Brown EN. Dynamic models of neural spiking activity. *Proc IEEE Conf Decision and Control.* 2007.

Czanner G, Eden UT, Wirth S, Yanike M, Suzuki W, Brown EN (2008) Analysis of within-and between-trial neural spiking dynamics. *J. Neurophysiol* In Press.

Daley D and Vere-Jones D (2003) *An Introduction to the Theory of Point Process.* 2nd ed., Springer-Verlag, New York.

Delgutte B (1980) Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. *J Acoust Soc Am* 68: 843-857.

Delgutte B, Hammond BM, Cariani PA (1998) Neural coding of the temporal envelope of speech: Relation to modulation transfer functions. In *Psychophysical and physiological advances in hearing,* AR Palmer, A Reese, AQ Summerfield, R Meddis (Eds). London: Whurr: 595-603.

Delgutte B, Kiang NYS (1984) Speech coding in the auditory nerve: I. Vowel-like sounds. *J Acoust Soc Am* 75:866-878.

Depireux, DA., Simon, JZ., Klein, DJ, Shamma, SA (2001) Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J Neurophysiol* 85: 1220-1234.

Dreyer AA, Czanner G, Eden UT, Lim HH, Anderson DJ, Brown EN (2007) Enhanced auditory neural threshold detection using a point process state-space model analysis. *COSYNE.*

Dreyer AA and Delgutte B (2006) Phase locking of auditory-nerve fibers to envelopes of high-frequency sounds: implications for sound localization. *J Neurophysiol* 96: 2327-2341.

Eden UT, Frank LM, Barbieri R, Solo V, Brown EN (2004) Dynamic analysis of neural encoding by point process adaptive filtering. *Neural Comput* 16: 971-998.

Eden UT, Truccolo W, Fellows MR, Donoghue JP and Brown EN (2004) Reconstruction of hand movement trajectories from a dynamic ensemble of spiking motor cortical neurons. *IEEE Engineering in Medicine and Biology Society* 2: 4017-4020.

Efron B (2004) The estimation of prediction error: Covariance penalties and cross-validation. *J Am Stat Assoc* 99: 619-632.

Escabi MA, Miller LM, Read HL, Schreiner CE (2003) Naturalistic auditory contrast improves spectrotemporal coding in the cat inferior colliculus. *J Neurosci* 23: 11489-11504.

Frank LM, Eden UT, Solo V, Wilson MA, Brown EN (2002) Contrasting patterns of receptive field plasticity in the hippocampus and the entorhinal cortex: an adaptive filtering approach. *J Neurosci* 22:3817-3830,

Gaumond PR, Molnar CE, Kim DO (1982) Stimulus and recovery dependence of cat cochlear nerve fiber spike discharge probability. *J Neurophys* 48: 856-873.

Geisler CD, Greenberg S (1986) A two-stage nonlinear cochlear model possesses automatic gain control. *J of Acoust Soc Am* 80:1359-1363.

Geisler CD (1998). *From sound to synapse: physiology of the mammalian ear.* New York: Oxford University Press.

Gill P, Theunissen F (2007) Spikes in the auditory forebrain: surprise, not intensity. *COSYNE.*

Goldberg JM, Brown PB (1969) Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *J Neurophysiol* 32: 613-636.

Gray PR (1967) Conditional probability analyses of the spike activity of single neurons. *Biophys J* 7: 759-777.

Greene WH (2007) *Econometric Analysis,* 6nd ed. Upper Saddle River, New Jersey: Prentice-Hall.

Grothe B (2003) New roles for synaptic inhibition in sound localization. *Nat Reviews Neurosci.* 4: 1-11.

Haslinger R, Andermann ML, Brown EN, Moore CI (2007) Point process models of neural response to inter-vibrissae interactions. *COSYNE.*

Heinz MG, Colburn HS, Carney LH (2001a) Evaluating auditory performance limits: I. one-parameter discrimination using a computation model for the auditory nerve. Neural Comput 13: 2273-2316.

Heinz MG, Colburn HS, Carney LH (2001b) Evaluating auditory performance limits: II. one-parameter discrimination with random-level variation. Neural Comput 13: 2317-2338.

Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Lio HH (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc Math Phys Eng Sci* 454: 903-995.

Hudspeth AJ, Corey DP (1977) Sensitivity, polarity, and conductance change in the response of vertebrate hair cells to controlled mechanical stimuli. *Proc Natl Acad Sci* 74: 2407-11.

Javel E, McGree JA, Horst W, Furley GR (1968) Temporal mechanisms in auditory stimulus coding. In GM Edelman, WE Gall and WM Cowan (Eds), *Auditory function: neurological bases of hearing* (pp.515-558). Wiley: New York.

Javel E (1980) Coding of AM tones in the chinchilla auditory nerve: Implications for the pitch of complex tones. *J Acoust Soc Am* 68: 133-146.

Jenison RL, Reale RA, Hind JE, Brugge JF (1998) Modeling of auditory spatial receptive fields with spherical approximation functions. *J Neurophysiol* 80: 2645-2656.

Johnson A, Kotz S (1970) *Distributions in Statistics: Continuous Univariate Distributions.* New York: Wiley.

Johnson DH (1980) The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *J Acoust Soc Am* 68: 1115-1122.

Johnson DH (1996) Point process models of single-neuron discharges. *J Comput Neurosci* 3: 275-299.

Johnson DH, Swami A (1983) The transmission of signals by auditory-nerve fiber discharge patterns. *J Acoust Soc Am* 4: 493-501.

Johnson DH, Tsuchitani C, Linebarger DA, Johnson MJ (1986) Application of a point process model to responses of cat lateral superior olive units to ipsilateral tones. *Hear Res* 21: 135-159.

Joris PX and Yin TCT (1995) Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences. *J Neurophysiol* 73: 1043-1062.

Kadia SC and Wang X (2003) Spectral integration in A1 of awake primates: neurons with single- and multipeaked tuning characteristics. *J Neurophysiol* 89: 1603-1622.

Kalluri S and Delgutte B (2003a) Mathematical models of cochlear nucleus onset neurons: I. Point neurons with many weak synaptic inputs. *J Comput Neurosci* 14: 71-90.

Kalluri S and Delgutte B (2003b) Mathematical Models of Cochlear Nucleus Onset Neurons: II. Model with Dynamic Spike-Blocking State *J Comput Neurosci* 14: 91-110.

Kandel ER, Schwartz JH, Jessell TM (2000) *Principles of Neural Science*. New York: McGraw-Hill Publishing Co.

Kass RE and Ventura VA (2001) Spike train probability model. *Neural Comput* 13: 1713-1720.

Kiang NYS (1966) Discharge patterns of single fibers in the cat's auditory nerve. Cambridge, Massachusetts: The MIT Press.

Kiang NYS, Moxon EC, Levine RA (1970) Auditory-nerve activity in cats with normal and abnormal cochleas. In: *Sensorineural Hearing Loss*, edited by GEW. Wolstenholme and J. Knight. London: Churchill, p. 241-273.

Kiang NYS, Moxon EC (1973) Tails of tuning curves of auditory-nerve fibers. *J Acoust Soc of Am* 54: 274-275.

Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J Neurophysiol* 60: 1799-1822.

Liang H, Bressler SL, Buffalo EA, Desimone R, Fries P (2005). Empirical mode decomposition of field potentials from macaque V4 in visual spatial attention. *Biol Cybern* 92: 380-392.

Liberman MC (1978) Auditory-nerve response from cats raised in a low-noise chamber. *J Acoust Soc of Am* 63: 442-455.

McKinney MF, Tramo MJ and Delgutte. B. Neural correlates of the dissonance of musical intervals in the inferior colliculus. Ph.D. Thesis. Massachusetts Institute of Technology, 2001.

Masterson B., Diamond IT, Harrison JM, Beecher MD (1967) Medial superior olive and sound localization. *Science* 155: 1696-1697.

McCullagh P, Nelder JA (1989) *Generalized Linear Models*. New York: Chapman & Hall/CRC.

Miller MI (1985) Algorithms for removing recovery-related distortion from auditory nerve discharge patterns. *J Acoust Soc Am* 77: 1452-1464.

Miller MI, Mark KE (1992) A statistical study of cochlear nerve discharge patterns in response to complex speech stimuli. *J Acoust Soc Am* 92: 202-209.

Narayan R. Grana G, Sen K (2006) Distinct time scales in cortical discrimination of natural sounds in songbirds. *J Neurophysiol* 96: 252-258.

Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point process. *J Am Stat Assoc* 83: 9-27.

Okatan M, Wilson MA, Brown EN (2005) Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity. *Neural Comput* 17: 1927-1961.

Paninski L (2004) Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Comput Neural Syst* 15: 243-262.

Pfeiffer RR (1966) Anteroventral cochlear nucleus: wave forms of extracellularly recorded spike potentials. *Science* 154: 667-668.

Qiu A, Schreiner CE, Escabi MA (2003) Gabor analysis of auditory midbrain receptive fields: spectro-temporal and binaural composition. *J Neurophysiol* 90: 456-476.

Quatieri TF (2001) *Discrete-time speech signal processing: principles and practice.* New York: Prentice Hall PTR.

Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W (1997) *Spikes: Exploring the neural code.* Cambridge, Massachusetts: The MIT Press.

Rhode WS, Smith P (1986) Encoding of timing and intensity in the ventral cochlear nucleus of the cat. *J Neurophysiol* 56: 261-286.

Rubinstein JT, Wilson BS, Finley CC, Abbas PJ (1999) Pseudospontaneous activity: stochastic independence of auditory nerve fibers with electrical stimulation. *Hear Res* 127: 108-118.

Sachs MB and Kiang NYS (1968) Two-tone inhibition in auditory-nerve fibers. *J Acoust Soc Am* 43: 1120-1128.

Sachs MB, Abbas PJ (1974) Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli. *J Acoust Soc Am* 56: 1835-1847.

Sachs MB, Young ED (1979) Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. *J Acoust Soc Am* 66: 470-479.

Sanes DH (1990) An in vitro analysis of sound localization mechanisms in the gerbil lateral superior olive. *J Neurosci* 10: 3494-3506.

Sanes DH, Rubel EW (1988) The ontogeny of inhibition and excitation in the gerbil lateral superior olive. *J Neurosci* 8: 682-700.

Sarma SV, Cheng M, Williams Z, Hu R, Eskandar E, Brown EN (2008) Characterizing and contrasting neural spiking activity in the sub-thalamic nucleus of Parkinson's patients to that of healthy primates. *J Neurosci* (submitted).

Schreiner CE, Langner G (1988) Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. *J Neurophysiol* 60: 1823-1840.

Sen K, Theunissen FE, Doupe AJ (2001) Feature analysis of natural sounds in the songbird auditory forebrain. *J. Neurophysiol* 86: 1445-1458.

Shinn-Cunningham BG, Best V, Dent ML, Gallun FJ, McClaine EM, Narayan R, Ozmeral E, Sen K (2006) Behavioral and neural identification of birdsong under several masking conditions. *Proc Int Symp Hear,* pp 164-170. Cloppenburg, Germany.

Siebert WM and Gray PR (1963) Random process model for firing patterns of single auditory neurons. *Q Prog Rep Res Lab of Elec MIT* 71: 241.

Siebert WM (1965) Some implications of the stochastic behavior of primary auditory neurons. *Kybernetik* 2: 206-215.

Sinex DG and Geisler CD (1981) Auditory-nerve fiber responses to frequency-modulated tones. *Hear Res* 4: 127-148.

Smith RL, Brachman ML (1980) Response modulation of auditory-nerve fibers by AM stimuli: effects of average intensity. *Hear Res* 2: 123-133.

Stevens KN (2000) *Acoustic phonentics.* Boston: The MIT Press.

Stone M (1977) An asymptotic equivalenence of choice of model by cross-validation and Akaike's criterion. *J R Stat Soc B* 39: 44-47.

Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci* 20: 2315-2331.

Theunissen FE, David SV, Signh NC, Hsu AH, Vinje WE, Gallant JL (2000) Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network: Comput Neural Syst* 12: 289-316.

Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown, EN (2005) A point process framework for relating neural spiking activity to spiking history, neural ensemble and extrinsic covariate effects. *J Neurophysiol* 93: 1074-1089.

Van Rossum MC (2001) A novel spike distance. *Neural Comput* 13: 751-763.

Victor JD, Purpura KP (1997) Metric-space analysis of spike-trains: theory, algorithms and application. *Network: Comput Neural Sys* 8: 127-164.

Wang L, Narayan R, Grana G, Shamir M, Sen K (2007) cortical discrimination of complex natural stimuli: can single neurons match behavior? *J Neurosci* 27: 582-589.

Yang X, Wang K, Shamma SA (1992) Auditory representations of acoustic signals. *IEEE Trans on inf theory* 38: 824-849.

Yates GK (1987) Dynamic effects in the input/output relationship of auditory nerve. *Hear Res* 27: 221-230.

Yin TC and Chan JC (1990) Interaural time sensitivity in medial superior olive of cat. *J Neurophysiol* 64: 465-488.

Young ED, Sachs MB (1979) Representation of steady-state vowels in the temporal aspect of the discharge patterns of populations of auditory-nerve fibers. *J Acoust Soc of Am* 66: 1381-1403.

Zacksenhouse M, Johnson DH, Williams J, Tsuchitani C (1998) Single-neuron modeling of LSO unit response. *J Neurophysiol* 79: 3098-3110.

Zhang X, Heinz MG, Bruce IC, Carney LH (2001) A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. *J Acoust Soc Am* 109: 648-670.

Chapter 5

CONCLUSION

1. *Summary of results*

The primary goal of this thesis was to explore models for auditory neural and behavioral data in the context of a likelihood and Bayesian framework. The use of this framework facilitates more informative approaches to better characterization of neural and behavioral properties of the auditory system. In addition, both frameworks allow for the estimation of the joint density of the data and obviate the need for hypothesis testing and multiple comparisons necessary for current methods.

Using the likelihood framework, we formulated two models for neural data as a dynamic point process. In Chapter 2, we used the EM algorithm to estimate maximum likelihood model parameters and illustrated that this methodology leads to lower estimates of neural threshold in A1 than current methods based on firing-rate based models for electric stimulation in the cochlear nucleus. We also used the likelihood framework in Chapter 4 to estimate ML parameters in a GLM framework for responses to natural acoustic stimuli in the auditory nerve, cochlear nucleus and field L, and showed that our framework leads to estimates of discrimination performance on par with behavioral observations. In Chapter 3, we used the alternative Bayesian framework and conducted estimation using MCMC sampling to employ several models without rewriting the estimation algorithms. We illustrated that this framework is applicable for analyzing behavioral auditory data as a dynamic process and illustrated inference estimation for psychometric functions.

For all three modeling examples examined in this thesis, we employed a principled methodology for analyzing and interpreting the neural and behavioral data. The methodology included first a model description and estimation based on likelihood principles and system information, such as point process insights for neural data and insights into the dynamic nature of the behavioral data. We illustrated estimation strategies in the likelihood and Bayesian

framework, using three different estimation strategies appropriate for the problem at hand, including Expectation-Maximization algorithm in Chapter 2, MCMC techniques with Bayesian sampling in Chapter 3 and an ML solution to the GLM using iteratively reweighted least squares algorithm in Chapter 4. Second, we employed goodness-of-fit analysis including Kolmogorov-Smirnov plots and statistics and autocorrelation function, and principled model selection criteria using Akaike Information Criteria. Third, we illustrated inference techniques such as threshold estimation and used decoding and cross-validation paradigms to estimate discrimination performance and evaluate the models. This three-fold analysis methodology serves as a good framework for representing and assessing models of auditory data.

2. *Implication for prosthetic design and mechanistic model improvement*

In Chapters 2 and 3 of the thesis, we examined models of neural and behavioral responses to electrical stimulation in the inferior colliculus in the auditory midbrain. The data used in these chapters was collected to study the potential for developing auditory midbrain implants. The aim of these prosthetics is to produce more localized, frequency-specific stimulation with lower thresholds than with cochlear stimulation, in order to allow for an increased number of frequency channels (Lim and Anderson, 2006). Additionally, researchers hope that the midbrain implant can facilitate an increase in the dynamic range of the neural response so that stimulation can be provided over a larger number of level steps that are able to be discriminated from one another (Lim and Anderson, 2006; Lim and Anderson, 2007). Our analysis here shows that using the likelihood-based point process framework, estimated thresholds are lower than those predicted using current analysis and thus lower than those found originally by Lim and Anderson (2006) used to argue that midbrain thresholds are lower to those found using cochlear stimulation. Our methodology illustrated in Chapter 2 can also be used for estimation of other neural response properties such as dynamic range and latency, and preliminary analysis not shown in this thesis illustrated larger values of dynamic range than those found by Lim and Anderson (2006). Therefore, our methodology provides even more

powerful arguments for the use of midbrain stimulation. Of course, application of our methodology will likely reveal lower thresholds and a larger dynamic range in auditory nerve fibers under cochlear stimulation then previously considered. In both cases, our methodology has the potential for extending the number and range of levels of stimulation for neural prosthetics, potentially using lower current levels that may lead to less current artifact and current spread.

Further studies should be conducted to investigate both neural and behavioral response in the same animal with stimulation in several auditory centers and to estimate neural models and psychometric functions as we have shown in Chapters 2 and 3 to determine whether lower neural and behavioral thresholds can be estimated using our paradigm. Analysis of these same experiments of simultaneous neural and behavioral data within our framework can also be used to investigate the pattern of neural thresholds in A1 by varying the locations of activation in the inferior colliculus. Lim and Anderson (2006, 2007) suggest that more rostral activation leads to lower thresholds, but perceptual consequences of this activation have not been examined. Our framework presented here offers an analysis technique that could lead to improvements in the placement of neural prosthetics and insight into whether an optimal placement location exists to maximize perceptual benefit.

3. *Extending current models to capture neural circuits mediating behavioral discrimination*

In Chapter 4, we discussed that modeling responses of songbirds in field L to conspecific birdsongs lead to discrimination between the songs similar to that observed behaviorally. This high level of neural discrimination performance estimation has implications for understanding the anatomical connections between the nuclei mediating behavioral song discrimination performance. Wang et al. (2007) suggests that the performance of their spike classifier can be thought of as a schematic representation of the circuitry represented in the avian birdsong system to discriminate songs. In this scheme, a downstream neuron would

respond only to specific songs by changes in its neural response. Wang et al. (2007) suggests that neurons involved in song selectivity can involve those areas downstream from field L such as the HVC, interfacial nucleus of the nidopallium and caudal mesopallium (Wang et al., 2007). Since the mechanisms remain unclear, Wang suggests recording from the input and these readout areas simultaneously to explore the biological mechanisms underlying discrimination performance (Coleman and Mooney, 2004). Such a scheme would truly be enlightening to explore whether the songbird system does degrade the fine temporal information, as suggested by the Wang et al. model of exponentially weighting each spike and temporal smearing by the Heavyweight function, or whether both rate-based and temporal information is preserved with the downstream neurons acting more like maximum likelihood estimators as we suggest here.

If simultaneous recordings from several neural populations thought to be jointly involved in behavioral discrimination performance were possible, the models presented here could be augmented to capture network effects and combine information across multiple neurons. In Chapter 1 of the thesis, we illustrate how decoding analysis can be accomplished by combining likelihood functions across multiple, independent neurons. Our finding in that study suggests that the discrimination performance of the system follows the performance of the best neuron, an observation strongly consistent with the lower envelope principle (Siebert, 1965; Parker and Newsome, 1998). However, to capture the network effects of neurons representing the discrimination circuitry, the assumption of independent neurons that we made in Chapter 2 would no longer hold. Rather, the models presented in this chapter would need to be augmented to reflect how the past spiking of downstream neurons affects the spiking of upstream neurons. Following previous statistical methods for analysis of neuronal interconnections using ML principles (Borisyuk et al., 1985; Brillinger 1975, 1988, 1992; Chornoboy et al., 1988; Okatan et al., 2005), we suggest several models of interconnected neurons within the GLM framework. For instance, one model could reflect the hypothesis that the downstream cells function as independent units, and would therefore be represented by the

model we explored in the thesis, with a single trial CIF

$$\lambda_{downstream}(b\Delta|\theta,\gamma,H_{b\Delta}) = \exp\left(\sum_{r=1}^{R}\theta_r g_r(b\Delta)\right)\exp\left(\sum_{j=1}^{J}\gamma_j n_{b\Delta-j}\right). \qquad (5.1)$$

Upstream neurons would, in turn, augment their firing propensity based on the past spiking history of the downstream neuron

$$\lambda_{upstream}\left(b\Delta|\theta^{up},\gamma^{up},H_{b\Delta},H_{b\Delta}^{up}\right)$$
$$= \exp\left(\sum_{r=1}^{R}\theta_r^{up}g_r(b\Delta)\right)\exp\left(\sum_{j=1}^{J}\gamma_j n_{b\Delta-j}\right)\exp\left(\sum_{j^{up}=1}^{J^{up}}\gamma_j^{up}n_{b\Delta-j}^{up}\right), \qquad (5.2)$$

where $H_{b\Delta}^{up}$ represents the upstream neuron's own past spiking history, γ_j^{up} the effect of the upstream neuron's past spiking history on current spiking, and θ^{up} the effect of the decomposition performed by the cell on current spiking. In this model of the upstream cell, we assume that the cell is somehow performing the stimulus decomposition independently of downstream cells. One could imagine that the upstream cell does not do any processing of the stimulus independently, but rather acts as a conglomerate of the input of the C downstream cells. In that case, the CIF that would better capture the upstream neuron's response would be

$$\lambda_{upstream}\left(b\Delta|\bar{\theta},\bar{\gamma},\gamma^{up},\overline{H_{b\Delta}},H_{b\Delta}^{up}\right)$$
$$= \exp\left(\sum_{c=1}^{C}\sum_{r=1}^{R}\theta_{r,c}^{up}g_r(b\Delta)\right)\exp\left(\sum_{c=1}^{C}\sum_{j=1}^{J}\gamma_{j,c}n_{c,b\Delta-j}\right)\exp\left(\sum_{j^{up}=1}^{J^{up}}\gamma_{j^{up}}^{up}n_{b\Delta-j^{up}}^{up}\right), \qquad (5.3)$$

where $\overline{H_{b\Delta}}$ represents the matrix of past spiking histories for the C downstream neurons, $\gamma_{j,c}$ represents the effect of the C downstream neurons' past spiking history on the spiking history of the upstream neuron.. In actuality, a cell probably performs some aspects of the stimulus decomposition while acting as a conglomerate of the downstream neurons to some extent. Therefore, another potential model would be some combination of (5.3) and (5.2) , where the effect of the downstream neurons' past spiking history and the cells processing could be scaled

down by some coefficients β_1 and β_2, respectively:

$$
\begin{aligned}
\lambda_{upstream}&\left(b\Delta\middle|\bar{\theta},\bar{\gamma},\theta^{up},\gamma^{up},\overline{H_{b\Delta}},H^{up}_{b\Delta}\right)\\
&= \exp\left(\sum_{c=1}^{C}\sum_{r=1}^{R}\theta^{up}_{r,c}g_r(b\Delta)\right)\exp\left(\beta_1 + \sum_{r=1}^{R}\theta^{up}_r g\_r(b\Delta)\right)\\
&\ast\exp\left(\beta_2 + \sum_{c=1}^{C}\sum_{j=1}^{J}\gamma_{j,c}n_{c,b\Delta-j}\right)\exp\left(\sum_{j^{up}=1}^{J^{up}}\gamma^{up}_{j^{up}}n^{up}_{b\Delta-j^{up}}\right).
\end{aligned}
$$

$$(5.4)$$

Network models of point process data is a burgeoning and exciting field of research (e.g. Espinosa and Gerstein, 1988; Li et al., 1999; Okatan et al., 2005) that could shed critical insight into the functionality of the neural circuits with the interpretation of the estimated encoding parameters as well as by any changes in decoding performance using different models to represent the neural circuit.

REFENCENCES

Borisyuk GN, Borisyuk RM, Kirillov AB, Kovalenko EI, Kryukov VI (1985). A new statistical method for identifying interconnections between neuronal network elements. *Biol Cybern* 52: 301-306.

Brillinger DR (1975). Identification of point process systems. *Ann Prob* 3: 909-929.

Brillinger DR (1988) Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol Cyber* 59: 189-200.

Brillinger DR (1992) Nerve cell spike train data analysis: A progression of technique. *J Am Stat Assoc* 87: 260-271.

Chernoboy ES, Schramm LP, Karr AF (1988). Maximum likelihood identification of neuronal point process systems. *Biol Cybern* 59: 265-275.

Coleman MJ, Mooney R (2004) Synaptic transformations underlying highly selective auditory representations of learned birdsong. *J Neurosci* 24: 7251-7265.

Espinosa IE, Gerstein GL (1988). Cortical auditory neuron interactions during presentation of 3-tone sequences: Effective connectivity. *Brain Res* 450: 39-50.

Li Z, Morris KF, Baekey DM, Shannon R, Lindsey BG (1999) Multimodal medullary neurons and correlational linkages of the respiratory network. *J Neurophysiol* 82: 188-201.

Lim HH, Anderson DJ (2006) Auditory cortical responses to electrical stimulation of the inferior colliculus: Implications for an auditory midbrain implant. *J Neurophysiol* 96: 975-88.

Lim HH, Anderson DJ (2007). Spatially distinct functional output regions within the central nucleus of the inferior colliculus: Implications for an auditory midbrain implant. *J Neurosci* 27: 8733-8743.

Okatan M, Wilson MA, Brown EN (2005) Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity. *Neural Comput* 17: 1927-1961.

Parker AJ, Newsome WT (1998) Sense and the single neuron: probing the physiology of perception. *Annu Rev Neurosci* 21: 227:277.

Siebert WM (1965). Some implications of the stochastic behavior of primary auditory neurons. *Biol Cybern* 2: 206-215.

Wang L, Narayan R, Grana G, Shamir M, Sen K (2007) cortical discrimination of complex natural stimuli: can single neurons match behavior? *J Neurosci* 27: 582-589.