# Population Encoding and Temporal Sequence Processing in the Auditory System

Mike Schachter

October 29, 2013

## Contents

## 1 Introduction

This section provides some background on the auditory system, some aspects of functional anatomy, and temporal processing of sequences. It is not required reading for the Aims section.

## 1.1 Functional Neuroanatomy of the Auditory System

In this section we'll emphasize findings from the mammalian auditory system and contrast them with the avian auditory system where necessary. The main idea is that by the time acoustic information has reached the thalamus, it has been broken down into binaural, tonotopically organized and temporally precise information about frequency and onset time, a separate stream carrying spatial localization information, and integration with somatosensory input used for things like echo cancellation.

*The Cochlea*

The cochlea is a tube separated into three parallel fluid-filled partitions, the scala vestibuli, scala media, and scala tympani. A sheet of epithelial and neural tissue divides the scala tympani from the scala media, comprised of the basilar membrane and organ of Corti.

Sound, a pressure wave that propagates through the air, reaches the ear canal and enters the middle ear. There, it pushes a system of bones that act as a mechanical lever. One of the bones, the stapes, pushes and pulls the cochlea in response to external vibrations.

The oscillatory movement of the stapes initiates a traveling wave down the basilar membrane of the cochlea. The basilar membrane is tonotopically organized, in that the mechanical energy from the sound is distributed to a specific place along the length of the membrane according to it's frequency. The energy of high frequency oscillations is distributed at the beginning of the membrane and energy of lowest frequencies at the end (Kandel, Schwarts, and Jessell 2000).

The cochlea is not a feedforward sensory organ. Outer hair cells that line the basilar membrane receive feedback from higher order nuclei. Outer hair cells deform in shape in response to depolarization, and can make the basilar membrane vibrate. These low amplitude vibrations can be heard as ”otoacoustical emissions” using sensitive microphones. Feedback control of the basilar membrane allows active amplification of low-intensity sounds to take place.

Vibrations of the basilar membrane also cause inner hair cells to oscillate around their preferred frequency. Information from inner hair cells is collected by cochlear ganglion neurons, which outnumber inner hair cells tenfold. Each cochlear ganglion neuron has a well-defined frequency-intensity tuning curve. They respond to increasing intensity with increasing firing rate, and can fire up to 500Hz. Frequency is coded tonotopically in the auditory nerve, and also by the periodicity of firing.

*The Cochlear Nucleus*

Axons in the auditory portion of the cranial nerve, which contains axons from the cochlear ganglion cells, synapse onto cells in the medulla in a region called the cochlear nucleus. It has several identified cell types:

1. Stellate Cells: tonically spike around preferred frequency, dependent on intensity.
2. Bushy Cells: fire single action potentials at onset of sound.
3. Octopus Cells: fire at onset to broadband sound, very temporally precise.
4. Fusiform Cells: respond to a broad range of frequencies, thought to encode location information along vertical axis.

By the time sound reaches the cochlear nucleus, it has been broken down by frequency, intensity, and onset time.

*The Lateral Lemniscus*

The lateral lemniscus is an area of the medulla that recieves input from axons coming from the cochlear nucleus. It is comprised of to the superior olivary complex and inhibitory neurons of the dorsal, ventral, and medial lateral lemniscus nuclei.

The superior olive uses inter-aural differences to localize sounds. The medial superior olive spatially localizes sounds using inter-aural timing differences (ITDs), and the lateral superior olive spatially localizes sounds using inter-aural intensity differences (IIDs). ITDs are useful for low frequency sounds, while IIDs are better for high frequency sounds.

The function of the lateral lemniscus is not clear (wikipedia.org), but it is separated into three nuclei. The neurons in each nucleus are strongly GABA-ergic and are segragated by response properties. Each represents temporal or spectrotemporal information in a different way.

*The Inferior Colliculus*

Input from the lateral lemniscus projects to the inferior colliculus, which has three anatomical subdivisions. The inferior collicus also recieves input from the somatosensory system, auditory cortex and thalamus and basal ganglia. It's implicated in behaviors such as the startle response and vestibu-ocular reflex. It also contains information about spatial localization. The avian counterpart to the IC is MLd, which functionally is where tuning to spectro-temporal modulation can be found.

*The Auditory Thalamus*

The medial geniculate body is the auditory thalamus. Functionally, it can be split into two anatomically distinct pathways; the lemniscal and the non-lemniscal (Hu 2003). The non-lemniscal pathway projects to the limbic system and layer 1 of the cortex, while the lemniscal system projects to layer 4 of primary auditory cortex. Functionally, non-lemniscal neurons may be tuned for change detection, and also rapidly adapt in response to behavioral training that uses auditory cues. Lemniscal neurons retain tonotopy and sharp spectro-temporal tuning properties. The avian homologue of MGB is OV, which has not been well studied, but seems to have neurons with sharp spectro-temporal tuning properties (Amin, Gill, and Theunissen 2010).

*The Avian Auditory Forebrain*

The avian auditory forebrain, recieving input from the thalamus, is the homologue to mammalian auditory cortex. The auditory forebrain is organized laminarly, with the layers NCM, L3, L2, L1, and CM going caudal to rostral (Wang, Brzozowska-Prechtl, and Karten 2010). Axons from the thalamus synapse in layer L2 and input is recurrently mixed from there among the other layers.

An analysis of receptive fields of neurons in Field L showed tonotopy in regards to the bandwidth of frequency responsiveness and a broadening of temporal window when moving from input to output layers (Kim and Doupe 2011). As noted in the next sections, layers CM and NCM are sites of stimulus-specific adaptation and task related plasticity.

## 1.2 Evidence for Experience-based Neural Plasticity

Early sensory experience shapes behavior. For example, Han et al. 2007 shows that rat pups raised in a single-tone acoustic environment had problems discriminating between frequencies at and near that single-tone, but improved discrimination performance for frequencies further away from that tone.

Kover et al. 2013 reared rats in acoustic environments which differed in their spectro-temporal composition, and found that the spectral bandwidth and tuning curve width of receptive fields varied with the environment. Rat pups reared with a temporal sequence of tone pips that alternate between two categorically-constructed frequency bands had receptive fields whose spectral tuning typically occupied one or the other spectral categories.

Sounds used in associative learning tasks significantly affect neural encoding. In area NCM of the Starling, Thompson and Gentner 2010 showed that training a bird to recognize a song in a reward driven task induces a strong reduction in response to cue sounds for neurons in NCM, while robust responses to novel and unfamiliar ssounds stay the same.

In networks, the relationship between signal and noise correlation seems to be related to task relevance. Jeanne, Sharpee, and Gentner 2013 shows that in area CM of the Starling, neuron pairs responsive to task-relevant sounds that have positive correlations in signal tuning but negative noise correlations, enhancing the ability to represent and decode the stimulus. For novel or task-irrelevant stimuli, they find that the signal and noise correlations between neurons are positive, making the stimulus more difficult to decode.

## 1.3 Temporal Sequence Processing in the Auditory System

In the avian system, it seems generally clear that the time window of integration for auditory stimuli increases with distance from the thalamus in the forebrain (Kim and Doupe 2011). A significant observation of generic temporal processing comes from stimulus-specific adaptation experiments. These experiments are typically conducted using an oddball stimulus protocol, where a sound, typically a pure tone, called the ”standard” is repeated with high probability, and a ”deviant” stimuli is embedded in the sequence. In a variety of regions and species, the neural response to the standard stimulus, while response to the deviant stays robust.

Ulanovsky, Las, and Nelken 2003 show that neurons in auditory cortex A1 of cats exhibit stimulus-specific reduction in tuning curve response towards the standard stimulus and an increase in tuning curve response for deviates. The effect held for both frequency and amplitude deviants. Ulanovsky, Las, Farkas, et al. 2004 is a more detailed examination of SSA in A1 neurons and shows that they adapt to standards at several timescales ranging from 150ms to tens of seconds. Neurons had a higher probability of spike failure in response to the standard.

In the Zebra Finch, Gill et al. 2008 shows that linear models perform better when the stimulus is preprocessed to represent the spectro-temporal surprise at short time scales (5-10ms). Beckers and Gahr 2012 show using call stimuli that SSA occurs for secondary auditory areas (CM) but not for primary auditory areas (Field L). They show that the secondary areas exhibit strongly synchronous firing in response to deviant stimuli.

Neurons sensitive to the combination of song syllables have been found both in Field L and the vocalization sensorimotor nucleus HVC (Margoliash and Fortune 1992, Lewicki and Arthur 1996). A class of excitatory projection neurons in NCM have been found to be sensitive to certain syllables only when presented in the context of song, and not when isolated (Schneider and Woolley 2013). It's clear that temporal sequence processing occurs in the auditory system, but it's mechanisms and function remain to be eludicated.

# 2 Aims

## 2.1 Aim 1: Auditory Network Information and Dynamics

For decades we have studied the stimulus encoding properties of single neurons with receptive fields. How much more information is contained in the population response? Can a distinction be made between intrinsic network dynamics and stimulus-driven activity? In this aim we will apply stimulus encoding and decoding techniques to try and answer this question. An emphasis will be placed on studying where, when, how and why do encoding and decoding models fail, and relating the time-varying performance of these models above to anatomical region, hemisphere, stimulus representation, and vocalization type.

1. How much of a neuron's spike train can be predicted from simultaneously recorded network activity? Compare performance of traditional receptive field models to models that encode a neuron's spike train from the simultaneously recorded spike trains of it's neighbors. Also fit models that predict a single neuron's spike train from the multi-electrode LFP.

2. What is the neural population code for sound stimuli? Relate the full population response of LFPs and spike trains to the stimulus using encoding and decoding models. Develop a time-varying pairwise coupling measure for the neural response to be used for improving model performance and increasing intuition about the neural population code.

## 2.2 Aim 2: Auditory Representation of Temporal Sequences

The ability to uniquely encode temporal sequences of sounds and discriminate between them is fundamental to human language. Like speech, Zebra Finch songs are nonrandom sequences of spectrotemporal elements, and neurons have been found in primary and sensorimotor auditory areas that are sensitive to syllable combinations (Margoliash and Fortune 1992, Lewicki and Arthur 1996). Such sensitivity, which can be described as a form of short-term memory, could be due to the intrinsic dynamics of the highly recurrent auditory network, and has not been rigorously studied at the population level (Buonomano and Maass 2009).

Human beings perceive phonemes in speech categorically, invariant to certain spectrotemporal variability. Swamp sparrows categorically perceive song syllables (Prather et al. 2009). Whether Zebra Finches perceive syllables categorically has not be studied, but they have been subjectively and algorithmically classified into a small number of categories that capture much of their spectrotemporal variability (Zann 1996, Lachlan, Verhagen, and Cate 2010). Also, syllable transitions and location within the song are not random and can be characterized using probabilistic models (Katahira, Okanoya, and Okada 2011).

In this aim we seek to determine what sub-networks in the brain are sensitive to syllable combination, and if syllable sequences are represented categorically.

1. Can a universal alphabet of Zebra Finch song syllables be built? Implement a probabilistic model of song syllables and their transitions so that syllables can be categorized and syllable sequence probabilities can be computed.

2. What is the best way to probe combination sensitivity, and where are combination sensitive subnetworks? Develop stimuli comprised of short sequences of syllables that test the encoding of a syllable with respect to the syllables that precede it. Perform multi-electrode dual hemisphere recordings from the auditory system and look for combination sensitive population responses in LFP and spike data. See section 3.1 for more information.

# 3 Experiments

The data for Aim 1 will be obtained from experiments that have already been run by Julie Elie in the Theunissen lab. Julie has performed dual hemisphere multi-electrode recordings from a variety of areas in the Zebra Finch forebrain. Her stimuli span the entire Zebra Finch vocalization repertoire, from calls to songs, and also include modulation-limited noise stimuli. The electrode tracts were labeled with dye and histology was performed. Spikes have been sorted and all the data is preprocessed and actively undergoing analysis.

The data for combination sensitivity will come from new experiments, designed to produce a rich dataset that minimizes the possibility of unpublishable results. Described briefly, juvenile and adult bird subjects will be housed next to a cage of ”familiar” male birds that provide a rich acoustic environment. The songs of the familiar birds will be recorded in isolation chambers. Another set of songs from another set of male birds that are not in contact with the subjects will be recorded and serve as unfamiliar songs.

On recording day, a subject bird will be anaesthetized and multi-electrode arrays will be implanted over the auditory cortex of each hemisphere. The stimulus protocol described in the next section will be played to the bird.

## 3.1 Stimulus Protocol

We want to test the combination-sensitivity of the neural response. To do so, we need to compare the response to an isolated syllable with the response to the syllable embedded into the context of other syllables. Each stimulus will be comprised of an isolated syllable, a sequence of 2-5 syllables, or an entire song. The syllables of any particular sequence will be from the same song, to mitigate confounds that could result from mixing up timbral identities of different birds. Stimuli will be repeated up to 5 times.

For a 5 syllable sequence, there are 5! = 120 different combinations to test, which is too many. The probablistic model we will fit allows us to compute a probability distribution over sequences. We will sample sequences from that distribution, making sure to include sequences that cover the range of probabilities.

Due to recurrence and adaptation effects, each stimulus may result in reverberating activity that continues in the silent period following it. We are actively seeking out this activity, and plan to leave on average 5 seconds of silence separating each stimulus presentation so that we can study it’s relation to the stimulus.

## 3.2 Recording of Song Standards

Each exemplar bird will be recorded in an isolation booth. Multiple renditions of each exemplar's songs will be recorded that encompass the natural variability in syllable length, inter-syllable spacing, and spectrotemporal properties (Glaze and Troyer 2006).

## 3.3 Bird Selection and Age of Exposure to Exemplars

We want to gather information on how both sex and age of exposure affect the prior probability of song sequences and thus the surprise of the deviants. The subjects will be evenly distributed among males and females. Two age groups will be examined for exposure to the exemplar cage; juveniles that are 60 days post-hatch, and adults that are at least 120 days post-hatch. Each subject will be exposed to the exemplar cage for at least one month preceding recording. We plan to have at least 5 subjects in each of the (gender,age) groups.

## 3.4 Electrophysiology

Recording will be done under urethane anestaesia. We will implant two 16-electrode arrays, one on each hemisphere. The arrays will be positioned over auditory regions, specifically targeting Field L, CMM, CLM, and NCM. The electrodes will be lowered in increments of 50-200$\mu$m. At each site, a search protocol will be executed to determine whether the cells are responsive to auditory stimuli. Once responsive cells are found at a site, the stimulus protocol (3.1) will be executed. When the protocol is finished, the electrodes will be lowered deeper to the next site.

Following the experiment, the bird will be overdosed with anesthesia, perfused, and debrained. The brain will be fixed in paraformaldehyde, frozen, and sliced. The electrode tracts will be identified and reconstructed using microscopy and digital image labeling, providing the anatomical regions of the recording electrodes.

# 4 Analysis

## 4.1 Stimulus Representation

The raw representation of sound is it's sound pressure waveform $x(t)$, typically sampled at 44.1kHz. We can safely assume that the signal can be represented as:

$$x(t) = A(t)cos(\theta(t)) \tag{1}$$

where $A(t)$ is slowly varying amplitude envelope, and $\theta(t)$ is a quickly changing phase. The amplitude envelope can be computed by low-pass filtering $x(t)$ and then rectifying.

An intuitive representation of $x(t)$ is the spectrogram, a time-frequency representation, written as $s(t, f)$. The spectrogram we typically use for analysis is formed from the short-time Fourier Transform of $x(t)$, with a Gaussian window of width 7ms, and an increment of 1ms. A logarithmic transformation is applied such that $s(t, f)$ is the power in decibels at time $t$ and frequency $f$.

The spectrogram analyzes a signal by the frequencies it contains, but one can go even further. The modulation power spectrum is the two-dimensional Fourier Transform of the spectrogram. It represents the temporal frequency and the spatial frequencies of the spectral frequencies present in a slice of spectrogram. The MPS, represented by $m(\omega_t, \omega_f)$, gives the power of the sound at temporal modulation frequency $\omega_t$ and spectral modulation frequency $\omega_f$ (Singh and Theunissen 2003).

Syllables with tight harmonic stacks will have power along the spectral frequency axis, and sounds with alot of temporal variation will have power along the temporal frequency axis. The MPS may be a good way of quantitatively capturing the "timbre" of a sound; the qualities of a sound that are not characterized by pitch. We can also compute a time-varying MPS, by windowing the spectrogram over time.

Finally, Zebra Finch vocalizations exist in several semantic classes, and each class is associated with a set of behaviors. Types of vocalizations include distance calls, used for long distance communication, tets, for short range communication, aggressive calls, thucks that signify impending takeoff, distress calls, and juvenile begging calls. The semantic type of a vocalization adds another dimension to the representation of the sound.

## 4.2 Representation of the Neural Response

The raw local field potential on an electrode is sampled at 380Hz, and we'll write it as $y(t)$. The LFP is a complex oscillatory signal comprised of many different frequencies. We will decompose the signal into a small set of $D$ narrowband components, and represent each component by it's time-varying amplitude and phase:

$$y(t) = Real\left[\sum_{k=1}^{D} z_j(t,k)\right] = Real\left[\sum_{k=1}^{D} A_j(t,k)e^{i\theta_j(t,k)}\right] \tag{2}$$

The function $z(t,k)$ represents the complex signal at time $t$ for component $k$. Each component has a characteristic time scale that is set by it's lowest observable frequency. The multi-electrode LFP we will work with is a complex-valued vector $\boldsymbol{z}(t,k) \in \mathbb{C}^n$, where $n$ is the number of electrodes, $t$ is time, and the signals are the $k$th bandpassed or decomposed components.

Unless otherwise specified, the spike trains will be represented by a long binary string, with a bin width of 1ms, with a value of 1 if a spike occurred in that bin. A population of spiking neurons will be represented by a binary vector $\boldsymbol{r}(t) \in \{0,1\}^c$, where $c$ is the number of neurons.

## 4.3 Encoding and Decoding Models

Generically, let $\boldsymbol{s}(t) \in \mathbb{R}^m$ be a time-varying stimulus, whether represented as an amplitude envelope ($m = 1$), slice of spectrogram ($m \approx 3600$), or time-varying modulation power spectrum. Let $\boldsymbol{z}(t)$ represent the population response, whether a set of LFPs, or spike trains, or both.

Encoding is the process of algorithmically finding a mapping from $\boldsymbol{s}(t)$ to $\boldsymbol{z}(t)$, which requires finding a linear filter that utilizes the history of $\boldsymbol{s}(t)$ to predict the instantaneous neural response $\boldsymbol{z}(t)$:

$$\hat{\boldsymbol{z}}(t) = F * \boldsymbol{s}(t) \tag{3}$$

The decoding model is in the other direction:

$$\hat{\boldsymbol{s}}(t) = G * \boldsymbol{z}(t) \tag{4}$$

Encoding and decoding are complementary processes. Naselaris et al. 2011 details five reasons for using encoding and decoding models, along with the efficacy of each type of model:

1. Does the neural response contain information about a specific type of feature representation? Both encoding and decoding models serve to answer this question, if features can be encoded to a neural response, then they should also be able to be decoded from the response.

2. Is the information represented in the neural response useful for behavior? Decoding models are more useful than encoding models in this instance, because downstream sensorimotor and motor areas must be decoding the relevant variables from the neural response.

3. Are there specific components of the neural response that contain more information than others about specific stimulus features? Prediction performance for both encoding and decoding models can be used for this purpose.

4. How much information does a neural response contain different types of feature representations? Encoding models answer this question best, because the prediction metric is based on the neural response, instead of diverse feature models.

5. What set of features provide a complete functional description of the neural response? Only encoding models can successfully do this. There is a multiplicity of feature representations that may be successfully decoded from the neural response, making decoding models less useful for this type of assessment.

## 4.4 A Time-varying Pairwise Coupling Graph Model of LFPs

When describing the coupling structure of LFPs or spikes, we can take at least two approaches. The naive approach is to compute time-varying couplings for each pair of neural responses, whether the pairs are comprised of two electrode LFPs or two spike trains.

### Local Field Potentials

We want to form a time-varying coupling graph from LFPs. However, local field potentials are complex oscillatory functions of time, comprised of the activity of thousands of neurons acting over multiple timescales. We want to understand coupling at multiple timescales, and in order to do this, we need to decompose the LFP into a set of components that are separated by their characteristic oscillatory frequencies.

*Decomposing the LFP into a Multi-timescale Signal with the Hilbert-Huang Transform*

Oscillations in mammalian brains have been well characterized by their location of production, associated behaviors, and frequencies of oscillation, through techinques such as EEG, electro- corticography (ECoG), and extracellular recordings using microelectrodes. Oscillations in different frequency bands serve different functional purposes. Oscillations in the Zebra Finch brain have not been well characterized, and our methods of analysis will try to avoid a-priori defined oscillation frequencies.

The Hilbert-Huang Transform (HHT) is a method for decomposing a nonstationary and nonlinear oscillatory function into a small set of ”intrinsic mode functions” (IMFs) that become progressively more narrowband (Huang et al. 1998). Each IMF has a well-defined Hilbert Transform, so that it can be successfully decomposed into instantaneous amplitude and phase components.

The HHT is well suited to LFPs, which are often nonstationary. We will use the HHT to decompose the LFP into narrowband signals that represent the LFP at different timescales. Low frequency IMFs, with frequency content less than 5Hz, contain information that varies on the timescale of one to several seconds. High frequency IMFs, with oscillations up to 200Hz, contain information that varies on a shorter timescale. We'll compare the IMFs to signals produced by bandpassing the LFP to have similar power spectrums.

*Phase Coupling Models*

Canolty et al. 2010 is a landmark paper that describes phase coupling between LFPs and spike trains. They elaborate on the existence of neurons that prefer certain multi-electrode LFP phase coupling structures. There are multiple phase coupling structures, and as such, multiple coactive cell assemblies. We are very aware of this paper and will follow up on their methods, examining the role of spike-triggered phase coupling with our data.

*The Pairwise Coherence Tensor*

A naive way to compute a time-varying pairwise coupling structure is to use a sliding window of size $\tau$ to compute the coherence between the LFPs of two different electrodes, $i$ and $j$. This gives us a time- frequency function for the correlation structure which we will call $\gamma_{ij\tau}(f,t)$. A value of $\gamma_{ij\tau}(f,t) = 0$ indicates no correlation between $i$ and $j$ at time $t$ and frequency $f$, while $\gamma_{ij\tau}(f,t) = 1$ is maximal correlation. We can integrate across the frequency bands of the coherence to produce the normal mutual information, a single number that summarizes the correlation betweeen the two components (Hsu, Borst, and Theunissen 2004):

$$I_{ij\tau}(t) = \int_0^{f_{max}} \log_2\left(1 - \gamma_{ij\tau}(f,t)\right)\, df \tag{5}$$

The choice of window size $\tau$ is dependent on the timescale of interest. We will decompose the LFP into a small set of narrowband components, by applying the HHT or by bandpass filtering. Each component has a characteristic timescale, dependent on the lowest frequency of interest. For example, if we're looking for 4Hz signals, and want to sample at least 20 cycles of the 4Hz component, we need a window size of at least 5 seconds. The higher the frequency, the smaller the window needed to adaquetely sample enough cycles of the relevant oscillations.

The pairwise coherence tensor is a 3-dimensional time-varying matrix, indexed by $I_{ij\tau}(t)$. It is a nonlinear second-order transformation of the neural response. Second order transformations are commonly performed in nonlinear system identification, such as in spike-triggered covariance, because some nonlinear systems can be better represent as linear transformations of nonlinear interactions between input features. The cost of this transformation is an increase in the dimensionality of the representation. With 32 electrodes and 5 different timescales, each time point is a 2544 element vector.

Stimulus encoding is a mapping between a time-varying stimulus vector $\boldsymbol{s}(t)$ and neural response. We will compute an encoding model by finding the weights of a multi-variate regression that maps $\boldsymbol{s}(t)$ to the coherence tensor $I(t)$. Stimulus decoding amounts to a multi-variate regression where the input signal is $I(t)$ and $\boldsymbol{s}(t)$.

## 4.5 A Probabilistic Model of Syllable Classes and their Transitions

### Song Segmentation and Syllable Representation

Due to the silent gap between syllables, it's relatively easy to segment a song by using a simple threshold on it's amplitude envelope. From there, we can compute the MPS of each syllable, and then compute a low-dimensional spectrotemporal representation $\boldsymbol{y}_i$ can be from the MPS', by applying PCA or another dimensionality reduction method. This produces an unlabeled dataset $\{\boldsymbol{y}_1, \boldsymbol{y}_2, ...\}$ of spectrotemporal representations.

### Universal Syllable Categories from a Hidden Markov Model

Given the dataset $\{\boldsymbol{y}_1, \boldsymbol{y}_2, ...\}$, we want to find a unique set of labels for each of the spectrotemporal representations that comprise an "alphabet" for Zebra Finch song. One way to do this is to fit a Hidden Markov Model to the dataset. Each $\boldsymbol{y}_i$ serves as an observation, and a fit HMM will provide the categorical label for it. HMMs also provide transition probabilities between the categorical syllable classes, and a model of spectrotemporal variability within each class.

## 4.6 A State Space Model of Combination Sensitivity

Traditional encoding models are likely to perform poorly for combination sensitive cells. A better model for combination-sensitive neural responses is the Kalman Filter, which has a continuous hidden state space as well as a continuous observation space. Once the data for combination sensitivity is obtained, we will experiment with fitting the neural response using such models.

# References

Amin, Noopur, Patrick Gill, and Frederic E. Theunissen (2010). "Role of the Zebra Finch Auditory Thalamus in Generating Complex Representations for Natural Sounds". In: *J. Neurophysiol* 104, pp. 784–798.

Beckers, Gabriel J. L. and Manfred Gahr (2012). "Large-Scale Synchronized Activity during Vocal Deviance Detection in the Zebra Finch Auditory Forebrain". In: *The Journal of Neuroscience* 32(31), pp. 10594–10608.

Buonomano, Dean V. and Wolfgang Maass (2009). "State-dependent computations: spatiotemporal processing in cortical networks". In: *Nature Reviews Neuroscience* 10, pp. 113–125.

Canolty, Ryan T. et al. (2010). "Oscillatory phase coupling coordinates anatomically dispersed functional cell assemblies". In: *Proceedings of the National Academy of Sciences* 107 no. 40.

Gill, Patrick et al. (2008). "What's That Sound? Auditory Area CLM Encodes Stimulus Surprise, Not Intensity or Intensity Changes". In: *Journal of Neurophysiology* 99, pp. 2809–2820.

Glaze, Christopher M. and Todd W. Troyer (2006). "Temporal Structure in Zebra Finch Song: Implications for Motor Coding". In: *The Journal of Neuroscience* 26(3), pp. 991–1005.

Han, Yoon K. et al. (2007). "Early experience impairs perceptual discrimination". In: *Nature Neuroscience* 10 No. 9, pp. 1191–1197.

Hsu, Anne, Alexander Borst, and Frederic E. Theunissen (2004). "Quantifying variability in neural responses and its application for the validation of model predictions." In: *Network: Computation in Neural Systems* 15, pp. 91–109.

Huang, Norden E. et al. (1998). "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis." In: *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 454.1971, pp. 903–995.

Hu, B. (2003). "Functional organization of lemniscal and nonlemniscal auditory thalamus". In: *Exp Brain Res* 153, pp. 543–549.

Jeanne, James M., Tatyana O. Sharpee, and Timothy Q. Gentner (2013). "Associative Learning Enhances Population Coding by Inverting Interneuronal Correlation Patterns". In: *Neuron* 78, pp. 352–363.

Kandel, Eric R., James H. Schwarts, and Thomas M. Jessell, eds. (2000). *Principles of Neural Science*. McGraw-Hill.

Katahira, Kentaro, Kenta Suzukiand Kazuo Okanoya, and Masato Okada (2011). "Complex Sequencing Rules of Birdsong Can be Explained by Simple Hidden Markov Processes". In: *PLoS One* 6(9).

Kim, Gunsoo and Allison Doupe (2011). "Organized Representation of Spectrotemporal Features in Songbird Auditory Forebrain". In: *The Journal of Neuroscience* 31(47), pp. 16977–16990.

Kover, Hania et al. (2013). "Perceptual and Neuronal Boundary Learned from Higher-Order Stimulus Probabilities". In: *The Journal of Neuroscience* 33(8), pp. 3699–3705.

Lachlan, R. F., L. Verhagen, and S. PetS. Peters. ten Cate (2010). "Are There Species-Universal Categories in Bird Song Phonology and Syntax? A Comparative Study of Chaffinches, Zebra Finches, and Swamp Sparrows". In: *Journal of Comparative Psychology* 124 no 1, pp. 92–108.

Lewicki, Michael S. and Benjamin J. Arthur (1996). "Hierarchical Organization of Auditory Temporal Context Sensitivity". In: *The Journal of Neuroscience* 16(21), pp. 6987–6998.

Margoliash, Daniel and Eric S. Fortune (1992). "Temporal and Harmonic Combination-Sensitive Neurons in the Zebra Finch's HVc". In: *The Journal of Neuroscience* 12(11), pp. 4309–4326.

Naselaris, Thomas et al. (2011). "Encoding and decoding in fMRI". In: *NeuroImage* 56, pp. 400–410.

Prather, Jonathan F. et al. (2009). "Neural correlates of categorical perception in learned vocal communication". In: *Nature Neuroscience* 12 no 2, pp. 221–228.

Schneider, David M. and Sarah M. N. Woolley (2013). “Sparse and Background-Invariant Coding of Vocalizations in Auditory Scenes”. In: *Neuron* 79, pp. 141–152.

Singh, Nandini C. and Frederic E. Theunissen (2003). “Modulation spectra of natural sounds and ethological theories of auditory processing”. In: *J. Acoustic. Soc. Am.* 114 (6), pp. 3394–3411.

Thompson, Jason V. and Timothy Q. Gentner (2010). “Song Recognition Learning and Stimulus-Specific Weakening of Neural Responses in the Avian Auditory Forebrain”. In: *Journal of Neurophysiology* 103, pp. 1785–1797.

Ulanovsky, Nachum, Liora Las, Dina Farkas, et al. (2004). “Multiple Time Scales of Adaptation in Auditory Cortex Neurons”. In: *The Journal of Neuroscience* 24(46), pp. 10440–10453.

Ulanovsky, Nachum, Liora Las, and Israel Nelken (2003). “Processing of low-probability sounds by cortical neurons”. In: *Nature Neuroscience* 6 no 4, pp. 391–398.

Wang, Yuan, Angnieszka Brzozowska-Prechtl, and Harvey J. Karten (2010). “Laminar and columnar auditory cortex in avian brain”. In: *Proceedings of the National Academy of Sciences* 107.28, pp. 12676–12681.

Zann, Richard A. (1996). *The Zebra Finch: A Synthesis of Field and Laboratory Studies.* Oxford University Press.