# Investigating dynamic neural representations of learning/memory in auditory cortex by sound reconstruction

**Nasim Winchester Vahidi**
Department of Electrical Engineering
University of California, San Diego
La Jolla, CA 92093
*nvahidi@eng.ucsd.edu*

**Xi Jiang**
Department of Neurosciences
University of California, San Diego
La Jolla, CA 92093
*x4jiang@ucsd.edu*

## Abstract

Songbirds rely on auditory processing of natural communication signals for the rare behavior of vocal learning—the ability to reproduce and recognize vocalizations through an adult model [1,4]. For this project, we obtained depth electrode recordings of local field potential from the caudal medial nidopallium, the songbird equivalent of auditory cortex. Anesthetized songbirds were given novel and familiar auditory objects, and the neural responses elicited by each object were used to predict the auditory objects provided, via a MATLAB-based reconstruction model. [2,3]. We expected that, as the novel objects were repeatedly presented, neural responses to repeats from the later part of stimuli presentation periods would produce better reconstructions (i.e. more similar to the actual objects) than the responses to earlier repeats. The results are inconsistent across birds, with a greater number of repeated presentations associated with worse reconstructions, in support of previously observed passive learning-induced changes in the stimuli-responsive neural networks.

## 1 Introduction

The songbird auditory system, being similar to the auditory system in humans, is well suited to studying changes in auditory encoding (Figure 2) [1]. Songbirds are notable for their complex vocally-mediated social interactions. These social interactions include mate selection/bonding, territory disputes, and individual vocal recognition, which requires forming an association between a particular individual and that individual's song [7,8]. As the birds begin to learn conspecific song motifs, one might expect the neural representation of auditory objects to form more readily while learning progresses. Indeed, associative learning has been shown to alter immediate early gene expression in songbird (canary) forebrains [9], with changes in the auditory telencephalon being best predicted by the pre-learning novelty of the stimuli. In European starlings, population coding was enhanced after associative auditory learning: after a binary choice associative learning task, where birds derived award directions from the order of natural song motifs given within stimuli pairs, stimulus-specific changes of the pattern of interneuronal correlations occurred within the songbird equivalent of auditory cortex [10].

In this study, auditory objects unfamiliar to the birds (in the form of starling song snippets), as well as objects that the birds had already heard more than 2,000 times during previous associative learning tasks, were presented repeatedly as single stimuli or double stimuli pairs (back-to-back snippets) to European starlings under anesthesia. Local field potential (LFP) recordings were collected via depth electrodes from the birds' caudal medial nidopallium (NCM), known to be the likely equivalent of mammalian auditory cortex in songbirds [1]. Since LFP represents a focal measure of neural population responses near each electrode contact, it is possible to reconstruct

various auditory stimuli based on correlative measures, such as a linear mapping between the neural response and the stimulus spectrogram [2,3]. Such reconstructions can be evaluated for how accurately they represent the original stimulus, i.e. the degree to which neural responses capture the information contained within the stimuli presented. Based on the known changes induced by novel auditory stimuli, we postulate that a similar progression may be observed in passive learning under anesthesia, whereby changes in neural population encoding in the NCM may be reflected as improvements in auditory object reconstructions after repeated presentations of novel stimuli.

Figure 1: European Starling (Sturnus vulgaris)

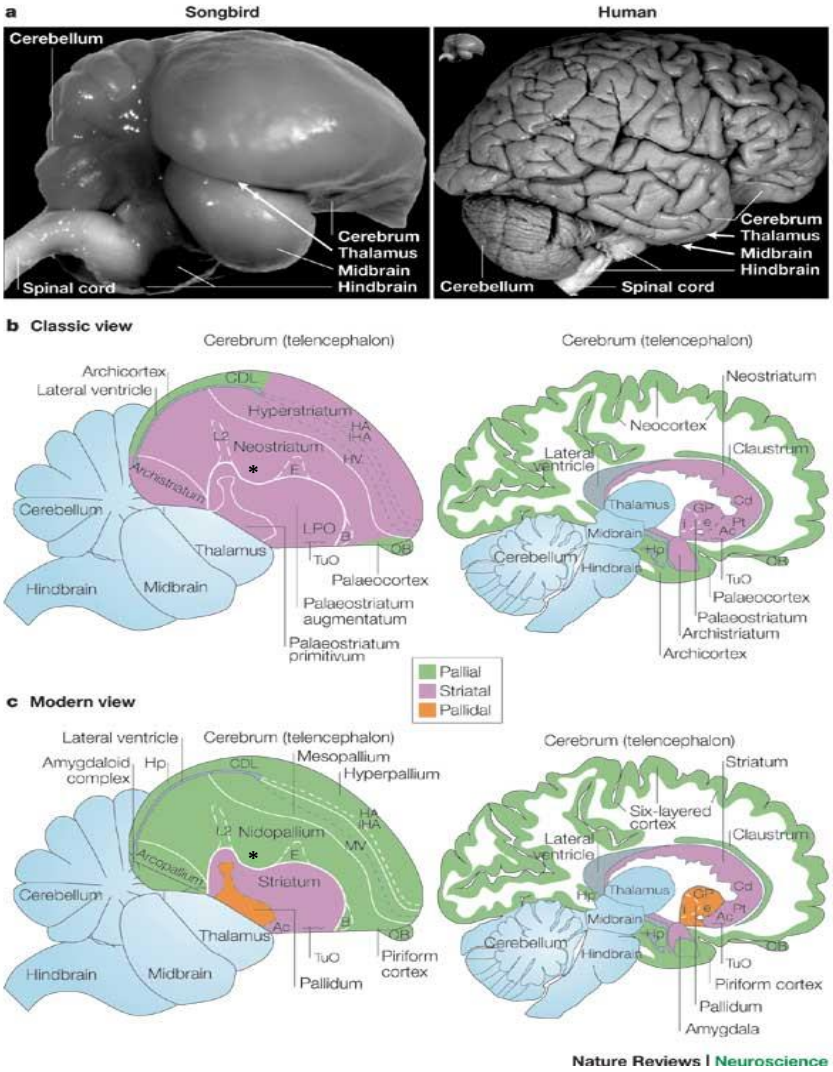

Figure 2: Comparative anatomy of human brain (right) and songbird brain (left) [1]. ***** marks the NCM.

# 2 Methods

## 2.1 Recording

Prior to performing neural recordings, subjects (two European starlings) were anesthetized with isoflurane (1.0–2.0% concentration in oxygen) and head-fixed. The upper layer of skull & trabecular were removed above NCM. A headpin was affixed to the skull caudal to the craniotomy with dental acrylic.

For the neural recordings, subjects were anesthetized with urethane (20% by volume, 7ml/kg) and head fixed in a stereotactic apparatus inside of a sound-attenuating chamber. Craniotomy was performed above NCM. A linear 32-channel silicon probe with 177um pads (Neuronexus) was coated with Di-I for later histological localization and advanced through the dura until single unit activity was visible on one or more channels in response to samples of starling songs (Figure 3). The local field potential was sampled at 25kHz and filtered (low pass at 100Hz) for offline analysis.

Recordings were obtained for multiple blocks of stimuli presentations (~500-800 presentations per block) at a given recording site, subject to maintenance of good isolation. The probe was then advanced by ~500µm to yield new units, until the ventral portion of NCM was reached.

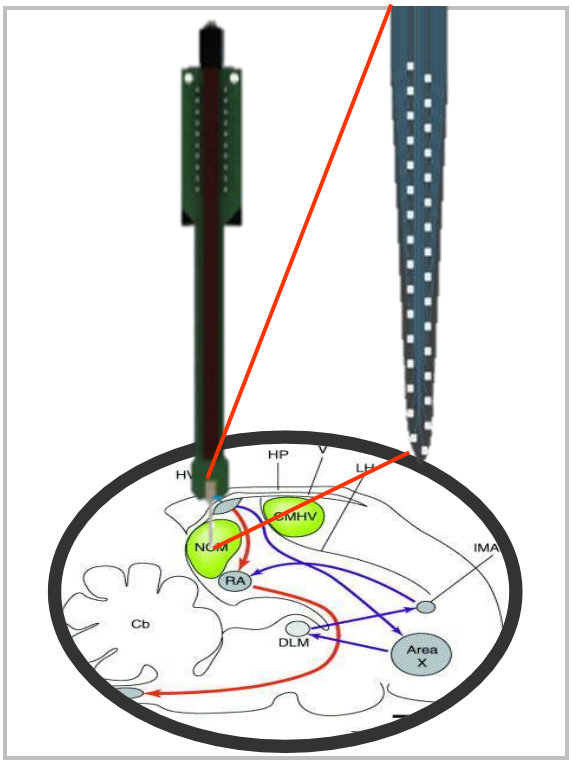


Figure 3: A linear 32-channel silicon probe was advanced through NCM.

## 2.2 Stimuli presentation

Two European Starlings birds were housed in a sound attenuating chamber (Acoustic Systems) with an operant panel. Prior to anesthesia, both birds received behavioral training on a standard GO/NOGO operant conditioning task[4]: 8 auditory objects (natural bird song motifs), denoted as **a-h**, were presented to the birds, with each presentation lasting 2.8-3.7s, containing either a single stimulus or a randomly ordered double-stimuli pair, with half the possible pairs indicating GO and

the other half indicating NOGO. Post-training, each bird had received more than 2,000 repeats of each stimulus, and stimuli **a**-**h**, therefore, can be termed "familiar".

Upon anesthesia, four novel/unfamiliar stimuli (**i**, **j**, **k**, **l**), alongside stimuli **a**-**h**, were presented to the birds in random order, either played in isolation or in sequential pairs (Figure 4). Four presentation periods/blocks were given to the birds at each recording site, with silent inter-motif intervals matched in duration to those found in the behavioral training. For each subject, the first blocks recorded at each site consisted of all familiar motifs in isolation, all GO pairs, and all NOGO pairs interleaved. Once isolation of putative units was maintained, further blocks would include all 144 pairwise combinations of novel and familiar motifs. Each ~3s presentation period was separated from another by at least two seconds of silence.

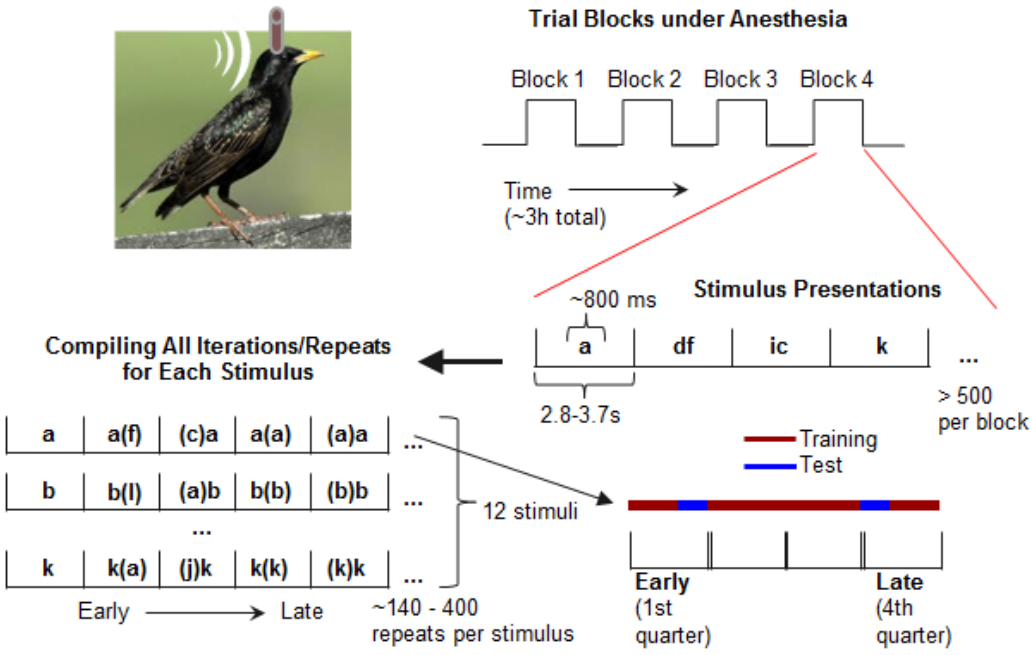


Figure 4: Stimuli presentation and data partitioning. Individual stimuli length is between 790 and 850 ms. The neural responses to a given stimulus within all time bins were compiled in temporal order, and the resulting data sets were partitioned into quarters, with 5 repeats in the Early/Late quarter being randomly chosen as the Early/Late test set, respectively.

### 2.3 Sound spectrum construction

To compute the power spectrogram of each song/stimulus (Figure 5), the time-continuous signal x(t) (where t $\in$ [-T, T] ) of the sound, through Fourier transform, was represented by X(f). The energy hidden in this signal was then calculated from $\int_{-\infty}^{\infty} x(t)^2 dt = \int |X(f)|^2 df$ (Parseval theorem), and the power (averaged energy) is given by lim T→∞ of (1/2T) $\int_{-\infty}^{\infty} |X(f)|^2 df$. The power spectrum was then calculated as G(f) = lim T→∞ of $(1/2T)|X(f)|^2$. 144 logarithmically spaced frequency bins (constructed via Constant Q transform over 6 octaves between ~200 Hz and ~12kHz) were employed to allow data manipulation in MATLAB.

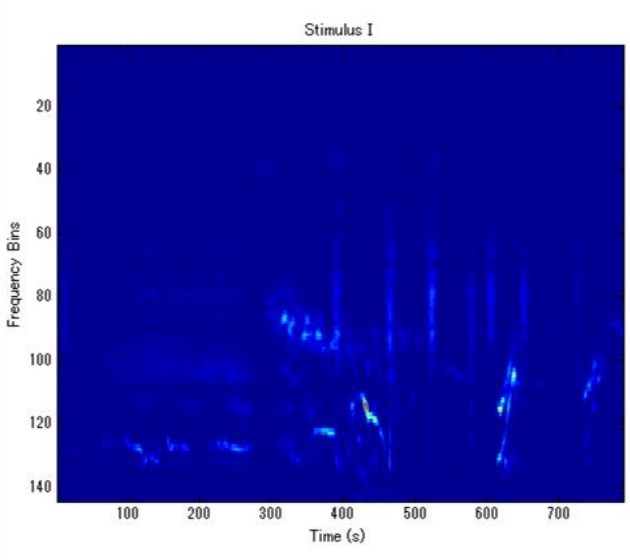


Figure 5. Top: Cartoon of stimulus I. Bottom: Power spectrum of stimulus I.

### 2.4 Reconstructing song spectrogram using optimal stimulus priors

In order to perform stimuli reconstruction, the optimal prior model [2,3,5] was implemented in MATLAB (Appendix). The model follows a linear mapping between the response of a population of neurons and the original stimulus: for a population of *N* neurons, response of neuron *n* can be represented at times $t_1, t_2 \ldots t_{max}$ (where $t_{max}$ = T) as $R(t, n)$. Because neurons in NCM are not phase-locked to the modulations in the original sound pressure waveform, the stimulus spectrogram $S(t, f)$ can be used to map linear stimulus response relationships by introducing a delay term $\tau$. The inverse function, $g(t, f, n)$, is a function that optimally maps $R(t, n)$ to $S(t, f)$, and the predicted reconstruction $\hat{S}(t, f)$ can be written as:

$$\hat{S}(t,f) = \sum_n \sum_r g(\tau, f, n) R(t-\tau, f) r_n(\tau)$$

For any given frequency window, $\hat{S}_f(t)$ from the neural population is independent of the other channels (estimated using a separate set of $g_f(t, n)$ )

$$\hat{S}(t) = \sum_n \sum_r g_f(\tau, n) R(t-\tau, n)$$

The function $g_f$ can be calculated by computing the reverse correlation: $g_f = C_{RR}^{-1} \; C_{RS_f}$ , where $C_{RR}$ and $C_{RSf}$ are the auto-correlation of neural responses and cross-correlation of stimulus and neural responses at different lags, respectively. The best fit model can be chosen by minimizing the mean-squared error between actual and reconstructed stimulus for a given frequency window:

$$\min e_f = \sum_n [(S_f(t) - \hat{S}_f(t)]^2$$

$$C_{RR} = RR^T, \quad C_{RS_f} = RS_f{}^T, \quad R = \begin{bmatrix} r_1(0) & r_1(1) & \dots r_1(\tau_{max}) \dots r_1(T) \\ 0 & 0 & \dots \; r_1(0) \dots r_1(T - \tau_{max}) \\ r_n(o) & r_n(1) & \dots \; r_1(\tau_{max}) \dots r_n \; (T) \\ 0 & 0 \quad \dots & r_1(0) \dots \; r_1(T - \tau_{max}) \\ & & \\ & & . \\ & & . \\ & & . \\ & & . \end{bmatrix}$$

$$S_f = [S(0, f) \dots S(T, f)]$$

The maximum time lag is set as $\tau_{max}$ = 100 ms. The entire reconstruction function is then described as the collection of functions for each spectral channel (Figure 5). The entire model can be defined, therefore, as $G = \{g_1, g_2 \dots g_F\}$, where $F$ = 144 from the Constant Q transform-derived stimulus frequency windows.

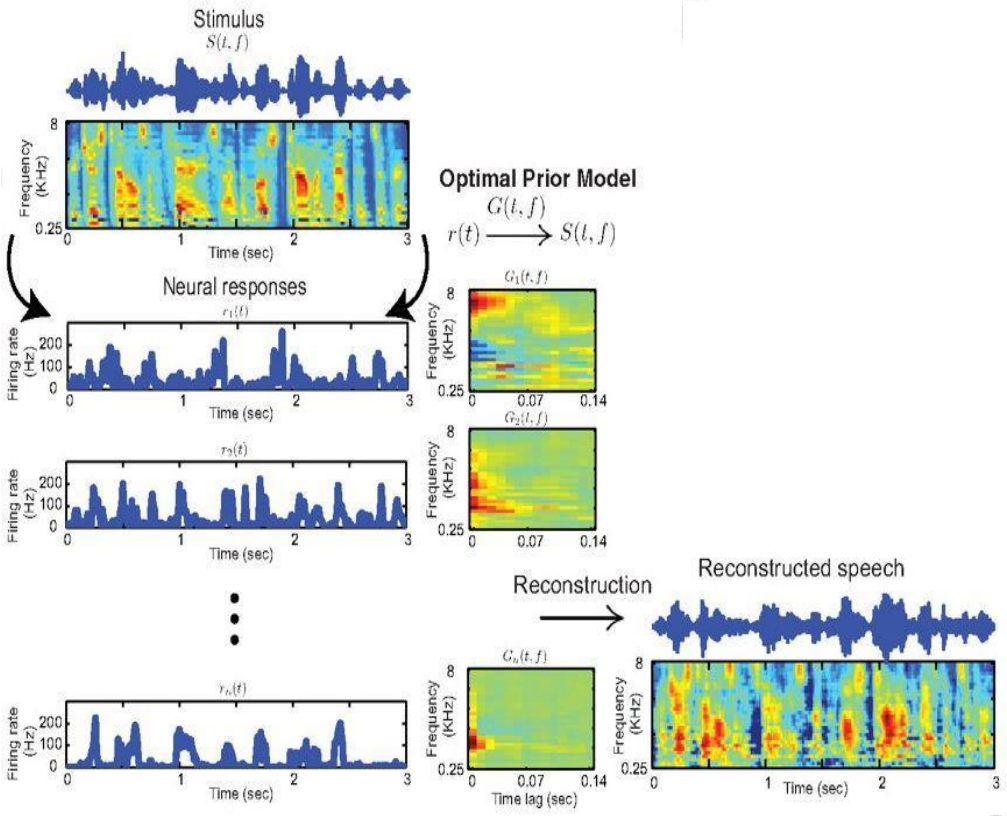


Figure 5. Optimal prior reconstruction (*G*) is the optimal linear mapping from a population of neuronal responses back to the sound spectrogram (*right*). Using optimal prior reconstruction, one can reconstruct the spectrogram of a sound: not only features that are explicitly coded by neurons, but also features that are correlated with them (left) [2].

# 3 Results

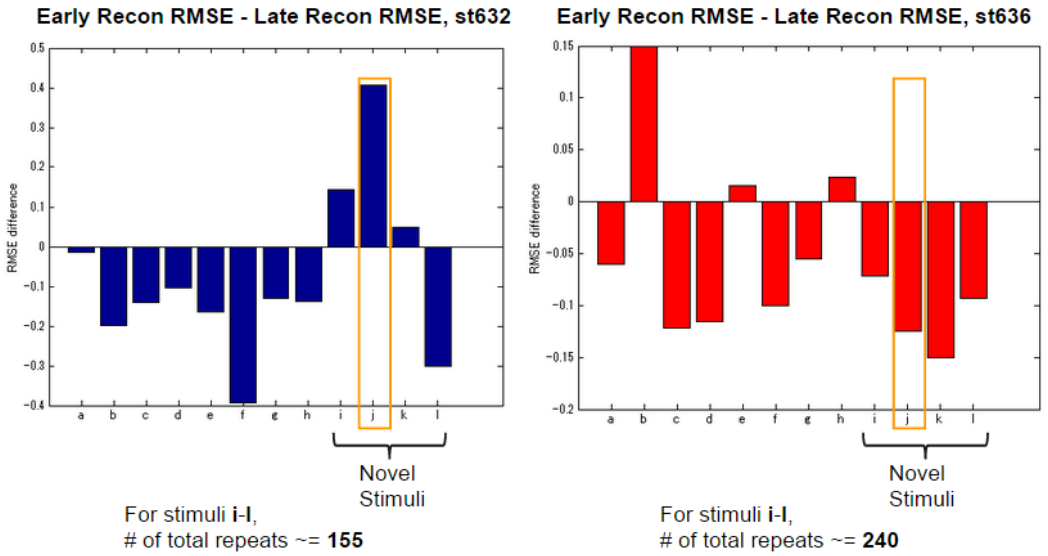


Figure 6. Reconstruction goodness-of-fit differences between early and late test sets for the starlings. While starling 632 showed improvement in 3 out of the 4 late sets for novel stimuli, starling 636 showed deficits uniformly. Orange square: stimulus **j**, shown in Figure 7.

Using the optimal prior model, reconstructions were obtained from the early and late test sets, as well as from all data in both test sets, for all 12 stimuli, based on the neural responses obtained from recordings at a single site in the right hemisphere NCMs (~1300-1500µm under the dura). Neural responses from bird st632, who experienced less novel stimuli repeats under anesthesia than bird st636 (about 80 repeats less for each novel stimulus), led to general improvements in 3 out of the 4 novel stimuli in terms of root mean square error (RMSE) differences between the early set reconstruction-to-original and the late set reconstruction-to-original errors (Figure 6, left). In contrast, responses to all familiar stimuli produced deficits in reconstruction over time.

Surprisingly, the neural responses from bird st636 led to deficits in late set reconstructions compared to the early sets (Figure 6, right) for all novel stimuli, and inconsistent changes for familiar stimuli. Across all frequency windows, the variance of correlations appeared to decrease over time for novel stimuli, such that the salient features in the stimulus becomes less distinguishable (Figure 7).

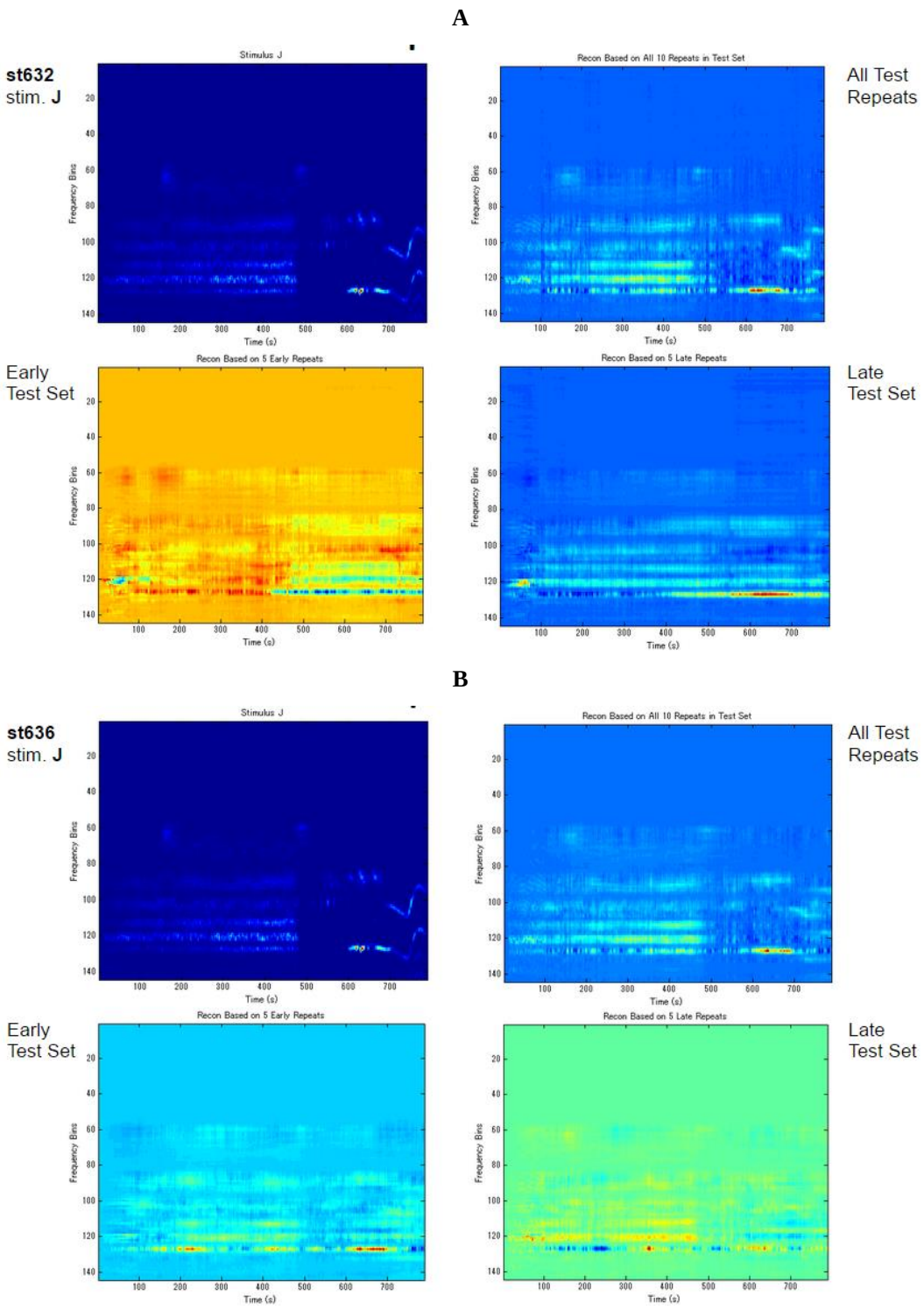


Figure 7. **A**: Stimulus J reconstructions from st632. **B**: Stimulus J reconstructions from st636. Features appear more salient in Early Test Set for st632, but in Late Test Set for st636. Early Test Set: reconstruction from the five randomly chosen repeats within the first quarter of all stimulus J repeats. Late Test Set: reconstruction from the five randomly chosen repeats within the last quarter of all stimulus J repeats. All Test Repeats: reconstruction from all ten Test Set repeats.

## 4 Discussions

Given the small sample size, the variabilities in late-learning reconstruction qualities for novel stimuli cannot be reliably determined. Nevertheless, it remains an interesting possibility to explore whether learning occurs in a "break-before-build" fashion, such that the initial improvements seen in st632 would be offset upon further repeats of stimuli presentations and lead to results similar to those from st636, prior to a consolidation period. It has been shown previously in songbirds that, during active learning, neural responses in NCM toward novel stimuli can decrease to a level beyond the baseline expected from adaptation [6]. A consolidation period, in both songbirds and humans, has been shown to rescue the deficits produced by learning and lead to task performance improvements that we normally associate with learning [11,12]. Therefore, the preliminary results we observed here might be indicating that passive learning affects neural responses in a manner similar to task-related active learning prior to memory consolidation, and further explorations via interleaving natural consolidation periods among stimuli presentations would be desirable in testing this hypothesis.


### Acknowledgments

We thank the members of the Gentner lab at UCSD for their continued support, and we are grateful for the wonderful starlings who lent their brains to us for purposes that will, sadly, remain beyond them.


### References


[1] Jarvis, E.D. *et al*. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews Neuroscience*, 6, 151-59.

[2] Mesgarani, N. *et al*. (2009). Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *Journal of Neurophysiology*, 102(6), 3329-39.

[3] Pasley, B.N. *et al*. (2012). Reconstructing speech from human auditory cortex. *PLoS Biology*, 10(1), e1001251.

[4] Gentner, T.Q. & Margoliash, D. (2003). Neuronal populations and single cells representing learned auditory objects. *Nature*, 424(6949), 669–74.

[5] Mesgarani, N. (2011). Robust representation of attended speech in human brain with implications for ASR. The Center for Language and Speech Processing (CLSP) Seminar. Accessed on Dec. 15th, 2014 from http://vimeo.com/45025398

[6] Knudsen, D.P. (2013). Learning to classify novel stimuli alters single neuron responses in the avian auditory cortex. Ph.D. Dissertation, University of California, San Diego

[7] Catchpole, C.K. & Slater, P.J.B. (1995). Bird song: biological themes and variations. University Press, Cambridge.

[8] Gentner, T.Q. & Hulse, S.H. (1998). Perceptual mechanisms for individual recognition in European starlings (Sturnus vulgaris). *Animal Behavior*, 56, 579-94.

[9] Jarvis, E.D. *et al*. (1995). Associative learning and stimulus novelty influence the song-induced expression of an immediate early gene in the canary forebrain. *Learning and Memory* 2: 62-80.

[10] Jeanne, J.M. *et al*. (2013). Associative learning enhances population coding by inverting interneuronal correlation patterns. *Neuron*, 78(2), 352-63.

[11] Shank, S.S. & Margoliash, D. (2009). Sleep and sensorimotor integration during early vocal learning in a songbird. *Nature*, 458, 73-7.

[12] Mednick, S. *et al*. (2003). Sleep-dependent learning: a nap is as good as a night. *Nature Neuroscience*, 6, 697-8.

## Appendix

```
%%%%% Building Stimuli Spectrum and Reconstruction%%%%%%%%%%%%%
% Finds the linear mapping between stimulus and neural data using optimal
% prior reconstruction, which accounts for any correlations in the stimuli.
% See Mesgarani et al. 2009, J. Neurophys. for more detail
ntrials = size(data,1);
blocksize = floor(ntrials/4);
[~,SortIndex] = sort(repeat_order);
data = data(SortIndex);

test_order = [randsample([1:blocksize],5),randsample([(blocksize*4-5):ntrials],5)];
train_set = setdiff([1:ntrials],test_order);
train_order = randsample(train_set,length(train_set));

sampleorder = [train_order test_order];
train = squeeze(mean(data(train_order,:,:),1));
all_test = data(test_order,:,:);
test = squeeze(mean(data(test_order,:,:),1));

CRR = train * train';
iCRR = pinv(CRR); clear CRR;
for sf = 1:size(stim,1)
    CRS = train * stim(sf,:)';
    g(sf,:) = iCRR*CRS;
    clear CRS
end

% now reconstruct on test set
rec = g*test;
crosscorr = corrcoef(rec',stim');
```