Learned context dependent categorical perception in a songbird

Tim Sainburg (tsainbur@ucsd.edu)

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

Marvin Thielk (mthielk@ucsd.edu)

Neurosciences Graduate Program, University of California San Diego

Timothy Q Gentner (tgentner@ucsd.edu)

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

Abstract

To distinguish between vocal elements in communication, both songbirds and humans rely on categorical perception (CP) of smoothly varying acoustic spaces. Boundaries in CP are often not fixed, and can be biased, in both human speech and birdsong, by context. How contextual information biases CP is not well understood. Here, we manipulate the information contained in auditory contextual cues to control CP of natural stimuli, using a 2-Alternative Choice (2AC) behavioral paradigm. We show that contextual information can bias CP of a smoothly varying acoustic stimuli, indicating that probabilistic inference contributes to context dependant perception. We follow this task with acute extracellular multichannel recordings on the secondary auditory nuclei caudal mesopallium (CM) and caudo-medial nidopallium (NCM) in the trained birds, investigating CP and contextual-related shifts in neural responses.

Keywords: context, categorical, perception, birdsong, vae

Introduction

Categorical perception (CP), the grouping of smoothly varying stimuli into discrete classes, plays an important role in organizing complex experiences into a shared representational space. The capacity to categorically represent stimuli allows for the abstraction of individual instances of a class to other instances. For example in speech each time we hear a novel exemplar of the word 'bat' we can draw upon our prior experiences of the word to make inferences about the referent, despite the variability in the acoustic signal underlying the phoneme /b/ between each utterance.

CP has been observed across sensory modalities (Etcoff & Magee, 1992; Eimas, Siqueland, Jusczyk, & Vigorito, 1971), and in numerous animal species (Wyttenbach, May, & Hoy, 1996; Baugh, Akre, & Ryan, 2008; Fischer, 1998; Nelson & Marler, 1989). In songbirds, categorical perception shows several parallels with human speech, such as population differences in categorical perception (Prather, Nowicki, Anderson, Peters, & Mooney, 2009) dependent upon prior experience (Thielk, Sainburg, Sharpee, & Gentner, 2017), as well as contextual dependencies influencing perceptual boundaries (Lachlan & Nowicki, 2015). In speech, for example, phoneme category perception is modulated by the position within a word (relative to other phonemes) (Marslen-Wilson & Welsh, 1978),

and word category perception is modulated by semantic context (Ganong, 1980). Recent investigations of songbird categorical perception have shown that similar to human speech, swamp sparrows modulate their CP of notes as a function of the position of the note in a syllable (Lachlan & Nowicki, 2015).

The computational and cognitive processes whereby contextual dependencies modulate CP are not well understood. To study these processes directly we developed a simple model to explain the interactions of stimuli and contextual cues. When a stimulus varies upon a single dimension x, the perceived value of x as a function of the true value of x and contextual information can be described by Bayes' rule:

$$\underbrace{P(x_{true} \mid x_{sensed}, cue)}_{\text{posterior}} \propto \underbrace{P(x_{sensed} \mid x_{true}, cue)}_{\text{likelihood}} \underbrace{P(x_{true} \mid cue)}_{\text{prior}}$$

Where *cue* refers to a contextual cue. By modulating the prior distribution of the categorical stimuli (x) with a cue, we predict that the perception of x will shift.
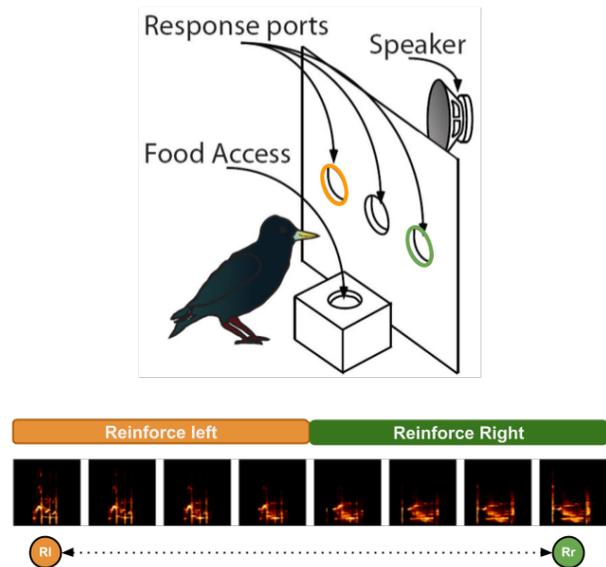

Figure 1: Experimental paradigm.

To implement this context-related CP shift in a natural stimulus environment, we created a two-alternative choice (2AC) category learning task using established operant conditioning

techniques (Figure 1 top) in which songbirds were trained to classify stimuli on a single dimension x, where x represents a smoothly varying syllable of birdsong generated from an interpolation in the latent space of a deep convolutional variational autoencoder (Figure 2) (Kingma & Welling, 2013). The stimuli generated from this interpolation (Figure 1, right) were split into two groups, with the first half of the interpolation being reinforced after pecking into the right response port and the stimuli generated from the left half of the interpolation being reinforced after a peck to the left response port (Figure 1 bottom).
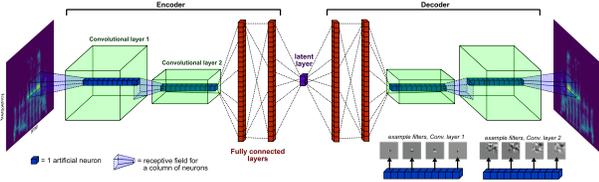

Figure 2: Generalized architecture of the convolutional variational autoencoder used. The number of neurons used in this experiment differs from the visualization.

Preceding each to-be-categorized target stimulus (x), we presented a cue stimulus, that provided predictive information about the category of the target stimulus (Figure 3). By treating this cue stimulus as a prior probability over x, we predicted that the determined posterior probability of x given sensory information and the cue stimulus would shift the classification of stimuli near the boundary between the two classes in the direction predicted by the cue stimulus.
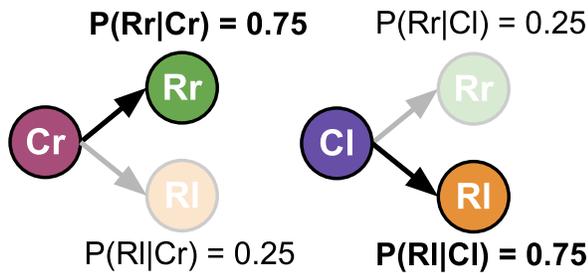

Figure 3: Contextual cues in experimental paradigm. Cue right (Cr) and Cue left (Cl) predict the reinforced right (Rr) and reinforced left (Rl) stimuli differentially.

Explicitly, we treat the likelihood of a target being sensed $P(x_{sensed} \mid x_{true}, cue)$ as a Gaussian probability distribution around the true target x_{true} as in Kording and Wolpert (2004):

$$P(x_{sensed} \mid x_{true}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_{true}-x_{sensed}}{\sigma_{sensed}}\right)^2}$$

and shift the prior probability as a function of the cue

$$P(x_{true} \mid cue) = \begin{cases} cue_{prob} & x_{true} \geq 62 \\ 1 - cue_{prob} & x_{true} < 62 \end{cases}$$

where cue_{prob} represents the predictive probability of the cue stimulus. We predict that birds will make a categorical decision based upon the posterior,

$$decision(x_{true}, x_{sensed}) = P(x_{true} \mid x_{sensed}, cue) category(x_{true})$$

where $category(x_{true})$ is simply the trained category label of x in the 2AC task:

$$category(x_{true}) = \begin{cases} 0 & x_{true} \geq 62 \\ 1 & x_{true} < 62 \end{cases}$$

Under this model, the categorical decision of the bird would be modulated by the prior, resulting in a shift in the inflection point along the stimulus dimension in the direction predicted by the cue (Figure 4 left), as opposed to a shift in the overall likelihood that all stimuli (regardless of their location along the stimulus dimension) will be categorized (Figure 4 right).
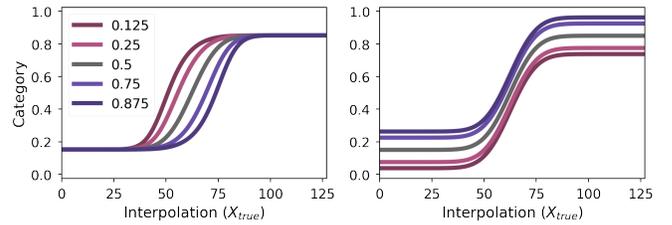

Figure 4: Additive (right) vs Bayesian integration (left) under different cue probabilities (color).

Results

We trained a total of 9 European starlings on our task, of which 7 birds learned the task to completion. One bird was discarded because of an insufficient amount of interaction with the operant conditioning device, and another was used for physiological investigations before reaching the context-dependency stage of the experiment.

Training paradigm

Birds were initially trained to differentiate between syllables generated via the two endpoints in a single interpolation. After several days of above-chance accuracy with one pair of interpolation endpoints, the number of interpolation endpoints was increased until the birds showed above accuracy classification of the endpoints of all 9 interpolations. After learning the correct response for endpoints in each interpolation, birds were transferred to the full stimulus set which included 127 stimuli (linearly spaced in latent space) spanning each of the 9 interpolations (1143 stimuli total). During this stage, the to-be-classified target stimuli were preceded by one of the two

cue stimuli (p=0.5), such that the cue provided no information about the following target stimulus.

After the birds were performing reliably above chance on each full interpolation stimulus set for several days, we altered the probabilities of the cue stimuli to provide context-dependant information. Initially, we set the predictive probability of each cue to its corresponding target to be 0.75. We then used the data from the next several thousand trials, to produce a psychometric function. If the inflection point (the 50% mark) of the fit psychometric function did not differ markedly between the left and right cue trials, the predictive probability was set to 0.875.

Psychometric Model

In each of the birds (n=7) we fit a psychometric (four parameter logistic) function both to the overall responses to stimuli in the left and right categories of the interpolation, as well as to each individual interpolation.
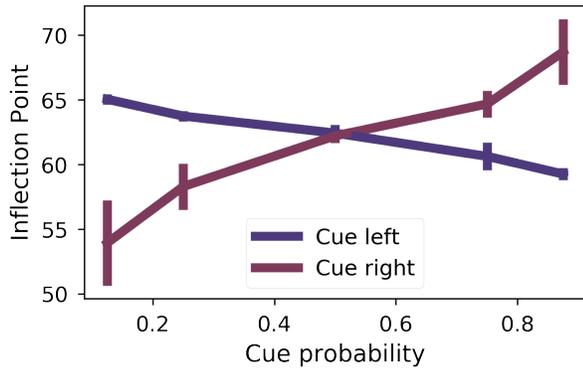
Figure 5: Inflection point of fit psychometric functions given cue.

We predicted a shift in the inflection point of the psychometric function (seen as a horizontal shift in figure 5 above), consistent with a shift in the location of the categorical boundary. The results in each condition support this prediction. Overall, the inflection point of each psychometric function was strongly correlated with the information provided by the cue (r(34) = 0.78, p < 0.001).

Bayesian Model

In addition to fitting a psychometric function capturing the shape of the behavioral responses, we fit a Bayesian model reflecting our probabilistic hypothesis described above. This model used five parameters: the shape of the Gaussian of the likelihood (σ_{sensed}), a parameter corresponding to side bias in the apparatus (γ), and parameters representing inattention to the cue stimulus (δ), the target stimulus (β), and overall inattention to the task (α).

$$bias_{side}(\gamma) = category(x_{true})(1 - 2(1 - \gamma)) + 1 - \gamma$$

$$likelihood = P(x_{sensed} \mid x_{true}, cue)(1 - \beta) + bias_{side}(\gamma)\beta$$

$$posterior \propto P(x_{true} \mid x_{sensed}, cue)(1 - \alpha) + bias_{side}(\gamma)\alpha$$

$$prior = P(x_{true} \mid cue)(1 - \delta) + bias_{side}(\gamma)\delta$$
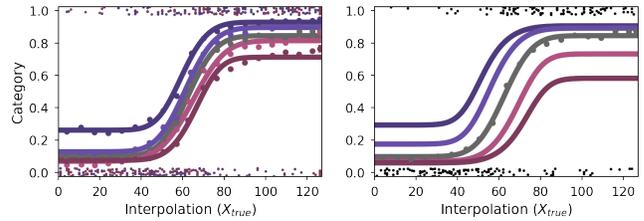
Figure 6: Bayesian prediction from fit of the model with cued priors (left) compared to that for flat priors on non-cued stimuli (right).

Ongoing work will explore how the Bayesian model parameters fit to a flat prior predicts individual and stimulus specific differences in biases to the categorical perception.

Physiological recordings

From trained subjects (currently n=3) and task-naive subjects (n=2) we recorded extracellular spiking activity in secondary auditory cortical regions NCM (caudo-medial nidopallium) and CM (caudal mesopallium) during playback of the categorical target and cue stimuli under light anesthesia (urethane) using 32 channel silicon probes. We spike-sorted (Chung et al., 2017) raw waveform data into putative single or multi-neuron sites, that were curated manually based upon waveform shape.
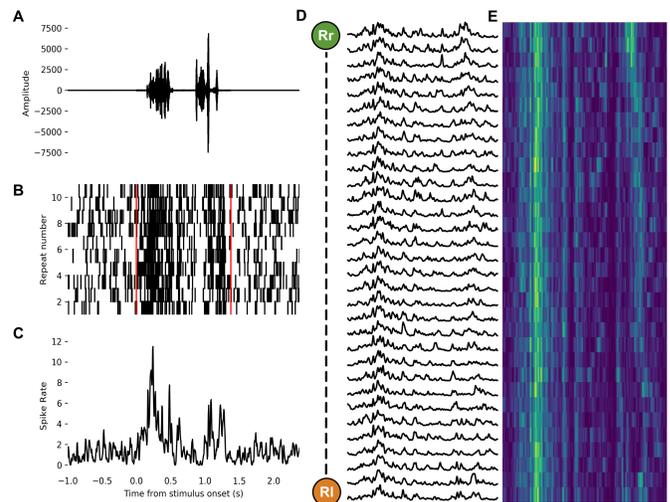
Figure 7: Overview of neural analysis. (A) audio from each trial are played back. (B) Raster of spiking data from a single neuron for the audio stimulus shown in (A). (C) Gaussian convolved PSTH of raster. (D) PSTH for each stimuli played back in on interpolation. (E) The same data as in D, shown as a heatmap.

Two playback conditions were used: one in which the cue preceded the target stimulus, and one in which the target stimulus was played back alone. In each condition, only 32 of the full 127 interpolation stimuli were presented during physiological recording due to time constraints. These stimuli were sampled more densely around the midpoints in the interpolations near the trained categorical boundaries.

For each stimulus and each sorted neuron, we computed a Peri-Stimulus Time Histogram (PSTH; bin size = 5ms) of each stimuli. We then convolved this PSTH with a Gaussian (σ = 5ms). This smoothed PSTH can then be used to compare single-neuron responses to individual stimuli as a function of the interpolation point (Figure 8). Ongoing work will explore within vs. between category similarity as well as context-dependent responses in these populations.
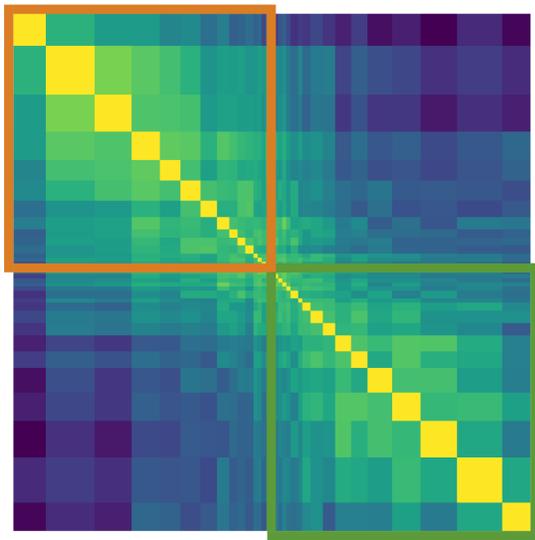

Figure 8: Correlations of neural activity concatenated across one population from CMM for one interpolation in latent space. Categories are outlined in orange and green.

Conclusions

Categorical perception involves a non-linear mapping between physical sensory stimuli and their representation in perceptual space, a phenomenon that appears to be fundamental to sensory integration. It has been observed across species, as well as across sensory modalities. This warping of perceptual space is not fixed. Contextual information can bias categorical perception, a phenomenon observed both in speech perception (Marslen-Wilson & Welsh, 1978) as well as in wild songbirds (Lachlan & Nowicki, 2015). Here, we explicitly trained songbirds on a categorical perception task, in which we actively modulated contextual information provided in the task. We found that using this paradigm, we were not only able to bias classification toward one category vs. the other, but also to shift the classification boundary between the two stimuli, a phenomenon which can be explained via probabilistic integration. Further analyses into this data will explore the physiological underpinnings behind the observed categorical perception, and its contextual bias.

References

Baugh, A. T., Akre, K. L., & Ryan, M. J. (2008). Categorical perception of a natural, multivariate signal: mating call recognition in túngara frogs. *Proceedings of the National Academy of Sciences*, *105*(26), 8985–8988.

Chung, J. E., Magland, J. F., Barnett, A. H., Tolosa, V. M., Tooker, A. C., Lee, K. Y., . . . Greengard, L. F. (2017). A fully automated approach to spike sorting. *Neuron*, *95*(6), 1381–1394.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–306.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, *44*(3), 227–240.

Fischer, J. (1998). Barbary macaques categorize shrill barks into two call types. *Animal Behaviour*, *55*(4), 799–807.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of experimental psychology: Human perception and performance*, *6*(1), 110.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244.

Lachlan, R. F., & Nowicki, S. (2015). Context-dependent categorical perception in a songbird. *Proceedings of the National Academy of Sciences*, *112*(6), 1892–1897.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive psychology*, *10*(1), 29–63.

Nelson, D. A., & Marler, P. (1989). Categorical perception of a natural stimulus continuum: birdsong. *Science*, *244*(4907), 976–978.

Prather, J. F., Nowicki, S., Anderson, R. C., Peters, S., & Mooney, R. (2009). Neural correlates of categorical perception in learned vocal communication. *Nature neuroscience*, *12*(2), 221.

Thielk, M., Sainburg, T., Sharpee, T., & Gentner, T. Q. (2017). Shared perceptual spaces for high-dimensional natural acoustic signals. In *Cosyne.* Salt Lake City, Utah.

Wyttenbach, R. A., May, M. L., & Hoy, R. R. (1996). Categorical perception of sound frequency by crickets. *Science*, *273*(5281), 1542–1544.