Laminar Processing in Primary Auditory Cortex ‘Untangles’ Vocalization Representations During an Active Categorization Task

Srivatsun Sadagopan (vatsun@pitt.edu)
Department of Neurobiology, University of Pittsburgh and the Center for the Neural Basis of Cognition
Pittsburgh, PA 15261, USA.

Manaswini Kar (mak433@pitt.edu)
Center for Neuroscience at the University of Pittsburgh and the Center for the Neural Basis of Cognition
Pittsburgh, PA 15261, USA.

Kayla Williams (kaw317@pitt.edu)
Department of Neurobiology, University of Pittsburgh
Pittsburgh, PA 15261, USA.

Vocalizations, such as animal calls and human speech, are produced with tremendous between-subject and inter-trial variability. A central function of auditory processing is to generalize over this variability and group calls into discrete categories. Previously, we developed an interpretable hierarchical model that accomplishes categorization by detecting features of intermediate complexity that capture the ‘gist’ of each call category. Neural responses selective for such features emerged in the superficial layers (L2/3) of primary auditory cortex (A1), and behavioral choices in a call categorization task were well-explained by this feature-based model. Here, we ask how call representations in different A1 laminae are modulated by task performance. We performed neural recordings from A1 of guinea pigs trained to categorize one conspecific call type from many other conspecific call types. Feature-selective neurons in A1 L2/3, while preserving their high selectivity to specific call features, showed increased output gain during task performance. Incorporating this increased output gain into our model resulted in better separated (‘untangled’) representations of categories. Together, these preliminary theoretical and experimental results reveal novel computational principles underlying auditory categorical representations and their modulation by attention. More broadly, our studies may provide insight into general computational principles underlying categorization across sensory modalities.

Keywords: categorization; generalization; vocalizations; auditory cortex; laminar processing; Neuropixels; electrophysiology; information theory; acoustic features.

Introduction

For vocal animals, including humans, conspecific vocalizations (‘calls’) are among the most ethologically relevant acoustic stimuli. Call recognition is computationally challenging because different within-species call types (conspecific calls) typically have highly overlapping spectral content and are produced with tremendous speaker-to-speaker and trial-to-trial variability (Hillenbrand et al., 1995; Wang, 2000). For accurate recognition, animals must generalize over this variability and group calls into behaviorally relevant categories. While categorization in the face of within-class variability has been extensively studied in the visual system, few theoretical frameworks exist for approaching these questions in a principled manner in the auditory system. Moreover, how the neural representation of sounds is transformed from one based on tuning for acoustic properties (e.g., spectral content) to one based on behaviorally relevant categories remains an open question. To address these questions, we developed a theoretical framework for call categorization (Liu et al., 2019; Sadagopan et al., 2023) broadly based on a face categorization algorithm (Ullman et al., 2002). The framework uses information-theoretic metrics and greedy search optimization to learn a set of contrastive intermediate-complexity features for each category (termed most informative features, or MIFs), that are likely to be present in most calls belonging to that category, and not present in outside-category calls. Such a theoretical model could achieve high categorization performance for classifying vocalizations in multiple species (Liu et al., 2019). In experiments using guinea pigs (GPs), highly social and vocal rodents with a rich repertoire of spectrotemporally complex calls, as an animal model, we showed that neurons with high selectivity for complex call features first emerged in the supragranular layers (L2/3) of A1 (Montes-Lourido et al., 2021). We also showed that the model, trained using only natural GP calls, could predict ~50% of the variance of the behavior of GPs performing a call categorization task using both natural and spectrotemporally modified calls (Kar et al., 2022).

Here, we asked how the activity of A1 neural populations that are tuned to spectral content or to informative features support the performance of a call categorization task. To do so, we simultaneously recorded the activity of ~100 neurons located across A1 laminae using implanted Neuropixels probes while GPs passively listened to calls and while GPs performed a call categorization task. Based on these empirical observations, we then extended our model to incorporate the effects of active listening during task performance. This allowed us to investigate the representational geometry of calls in different A1 laminae and their modulation by attention.

Methods

Behavioral task GPs performed a Go/No-go call categorization task. GPs moved to one side of a behavioral arena to initiate trials. A call stimulus, drawn from multiple examples (typically eight calls made by four unrelated GPs) belonging to a 'Go' call category or multiple examples drawn from several 'No-go' call categories, was presented from an overhead speaker. GPs were required to move to the other side of the arena upon hearing a 'Go' category stimulus to receive a food pellet reward. False alarms incurred a time-out.

Neural recordings During this freely-moving task, we obtained spiking and local field potential activity across A1 laminae using implanted Neuropixels probes. Single units were clustered using established techniques, and their laminar locations (supragranular, L2/3; granular, L4; infragranular, L5/6) assigned by registering spike locations to the current source density computed from local field potential responses to tones.

Results

Feature Selective A1 L2/3 Neurons Increased Output Gain During Active Listening Fig. 1A shows the post-stimulus time histogram of an example 'wheek' call selective neuron responding to various calls. By fitting responses using linear/nonlinear models in the passive and active listening conditions, we found that the neuron's feature preference remained stable across conditions (Fig. 1B). The output gain, however, was greatly increased during active listening (Fig. 1C).

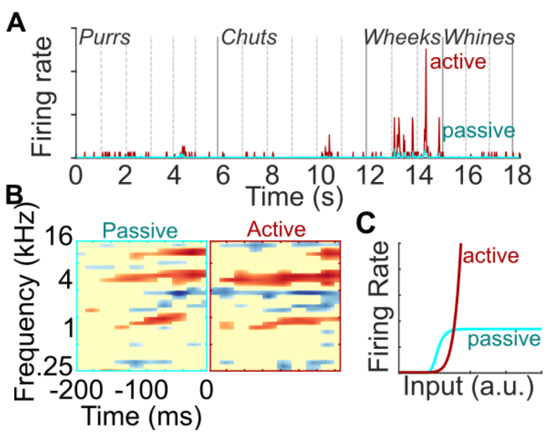

Figure 1 A1 L2/3 neurons preserved their selectivity but showed increased output gain during active listening.

Increased Output Gain Untangles Call Representations in A1 L2/3 In our model, calls are represented in terms of spectral content (cochleagram) in the input layer. We have previously shown that model input layer responses are consistent with the responses of A1 L4 neurons. In this spectral representation, call categories were highly overlapping when visualized in reduced dimensional space (Fig. 2B). From these inputs, the model learnt MIFs that were optimal for categorizing each call category from all other calls (Fig. 2A). The 'response' of each MIF to a call was taken to be the maximum value of the normalized cross-correlation function between the MIF and stimulus cochleagram. When MIF output rates were taken to be a linear function of the MIF responses (modeling the passive listening condition), MIF responses already showed better separation of call categories compared to the spectral representation (Fig. 2C). To model the active listening condition, we imposed a power law relationship between the MIF response and the MIF output rate (magenta line in Fig. 2D). This resulted in even better separated call category representations (Fig. 2E). We quantified the distances between the call categories using *d'* as a metric. Applying increased output gain to neurons tuned to spectral content did not result in better separated call category representations (Fig. 2F, gray). However, increasing the output gain of MIF-selective neurons consistently increased the distances between call categories (Fig. 2F, black).

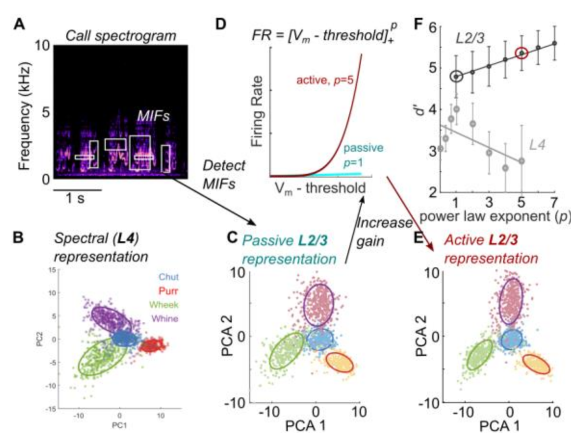

Figure 2 Untangling of call representations during active categorization task performance.

Conclusions

Our preliminary results from these ongoing experiments and simulations demonstrate how active listening helps 'untangle' call representations in A1 L2/3. This strategy of detecting contrastive features for categorization and increasing the output gains of detected features during attentive perception to increase categorical distances may be a general computational principle operating across sensory modalities.

References

Wang, X. (2000). On cortical coding of vocal communication sounds in primates. *Proceedings of the National Academy of Sciences*, 97(22), 11843-11849.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, 5(7), 682-687.

Liu, S. T., Montes-Lourido, P., Wang, X., & Sadagopan, S. (2019). Optimal features for auditory categorization. *Nature Communications*, 10(1), 1302.

Sadagopan, S., Kar, M., & Parida, S. (2023). Quantitative models of auditory cortical processing. *Hearing Research*, 429, 108697.

Montes-Lourido, P., Kar, M., David, S. V., & Sadagopan, S. (2021). Neuronal selectivity to complex vocalization features emerges in the superficial layers of primary auditory cortex. *PLoS Biology*, 19(6), e3001299.

Kar, M., Pernia, M., Williams, K., Parida, S., Schneider, N. A., McAndrew, M., ... & Sadagopan, S. (2022). Vocalization categorization behavior explained by a feature-based auditory categorization model. *Elife*, 11, e78278.