Neuronal Population Encoding of Identity in Primate Prefrontal Cortex

Sharma, KK; Diltz MA; Lincoln, T; Romanski, LM*

University of Rochester, School of Medicine and Dentistry, Department of Neuroscience, Rochester NY

*Correspondence to:
Lizabeth Romanski, Ph.D.

Lizabeth_romanksi@urmc.rochester.edu

Departments of Neuroscience, and Psychiatry
The Del Monte Institute for Neuroscience
University of Rochester Medical Center
601 Elmwood Avenue,
Rochester, New York 14642

Abstract

The ventrolateral prefrontal cortex (VLPFC) is consistently active during the perception of faces and voices. However, little is known about what categorical features of social stimuli drive neural activity in this region. Since perception of identity and expression are critical social functions, we examined whether neural responses to naturalistic stimuli were driven by these two categorial features in the prefrontal cortex. We recorded single neurons in the VLPFC, while two macaques viewed short audiovisual videos of unfamiliar conspecifics making expressions of aggressive, affiliative, and neutral valence. Of the 285 neurons responsive to the audiovisual stimuli, 111 neurons had a main effect (two-way ANOVA) of identity, expression or their interaction in their stimulus related firing rates; however, decoding of expression and identity using single unit firing rates rendered poor accuracy. Interestingly, when decoding from pseudopopulations of recorded neurons, the accuracy for both expression and identity increased with population size, suggesting that the population transmitted information relevant to both variables. Principal components analysis of mean population activity across time revealed that population responses to the same identity followed similar trajectories in the response space, facilitating segregation from other identities. Our results suggest that identity is a critical feature of social stimuli that dictates the structure of population activity in the VLPFC, during the perception of vocalizations and their corresponding facial expressions. These findings enhance our understanding of social behavior beyond the temporal lobe in macaques and humans.

Introduction

Facial and vocal expressions are a fundamental component of communication in humans and non-human primates (Blair, 2003; Darwin, 1965; Ekman & Friesen, 1971; Niedenthal & Brauer, 2012; Parr & Heintz, 2009; Partan, 2002). In macaques, viewing conspecifics and their expressions produces robust activity through the occipital, temporal, and ventrolateral prefrontal cortices (Russ & Leopold, 2015; Shepherd & Freiwald, 2018). The ventrolateral prefrontal cortex (VLPFC) exhibits a high degree of responsivity to social stimuli. It contains face cells (O Scalaidhe, 1999; O Scalaidhe et al., 1997) and face patches (Tsao, Schweers, et al., 2008), neurons responsive to the angle of a viewed face (Romanski & Diehl, 2011), and neurons selective for species-specific vocalizations (Hage & Nieder, 2015; Romanski et al., 2005; Romanski & Goldman-Rakic, 2002).

These findings suggest that the VLPFC may serve a critical social function. However, little is known about what social variables are represented by single neuron and population level activity in the VLPFC. This is a stark contrast to the face patch system in the temporal lobe (Tsao, Moeller, et al., 2008), where a variety of features that drive neural responses have been identified. Anterior face patches contain neurons with selectivity for identities, whereas middle face patch activity correlates with head orientation of the viewed stimulus (Freiwald & Tsao, 2010). Patches in the fundus of the superior temporal sulcus are sensitive to changes in expression (Taubert et al., 2020) while face patches outside the traditional network display joint encoding of multiple

variables, including expression and identity but also gaze and head orientation (Yang & Freiwald, 2021). Notably, information about identity is broadly distributed in the face patch network because it can be reliably decoded from small neural populations across many face patches based on representation of neurons as axes in a high-dimensional face space (Chang & Tsao, 2017). While neurons with differential responses for the angle of a viewed face have been identified in VLPFC (Romanski & Diehl, 2011), a systematic exploration of their selectivity for identity or type of expression has yet to be conducted.

A further complication in assessing the social function of the VLPFC arises from its role as a multisensory integrator. A VLPFC neuron's multisensory response to an audiovisual face-vocalization stimulus is frequently a non-linear combination of its component responses to the face or vocalization alone (Diehl & Romanski, 2014; Sugihara et al., 2006). While this has the advantage of drastically increasing the possible set of responses a neuron or population of neurons can have (Rigotti et al., 2013), it also illustrates how response characteristics to a single sensory modality can fail to represent ethologically valid stimuli. Simply put, non-linear multisensory integration provides direct evidence that conclusions drawn from neural responses to a static image or isolated vocalization will not scale to natural versions of these inherently multisensory stimuli.

To glean more insight into the categorical social features that drive VLPFC neural activity during social communication, we recorded single unit activity in the VLPFC while macaques perceived naturalistic expressions from unfamiliar conspecifics. Given

the primacy of identity and expression in similar studies (Gothard et al., 2007; Kuraoka et al., 2015; Kuraoka & Nakamura, 2007; Yang & Freiwald, 2021), we varied stimuli by conspecific identity and expression. We presented ethologically valid audiovisual movie clips from three different conspecifics, each making three discrete types of audiovisual expressions with distinct valances (aggressive, affiliative, and neutral, respectively; Hauser and Marler 1993; Partan 2002). We recorded neurons across the VLPFC and analyzed the subset that displayed consistent firing rate changes in response to at least one of the nine stimuli presented. We found that single neurons had low selectivity for single stimuli and even lower selectivity for expression or identity as a category. Rather, they responded to various stimuli over the stimulus period. Despite heterogenous firing patterns, both expression and identity could be accurately decoded from populations of neurons. Furthermore, we found that population responses to the same identity followed coherent trajectories over time that separated them from other identities in the population response space.

Methods

Subjects and surgical procedures

Extracellular recordings were performed in two rhesus monkeys (*Macaca mulatta*): Subject 1 (male: 12.0 kg, 7yo) and Subject 2 (male: 16.5 kg, 12yo). All procedures were in accordance with the National Institute of Health's *Guidelines for the Care and Use of Laboratory Animals* and the University of Rochester Care and Use of Animals in Research committee. For Subject 1, prior to recordings, a titanium head post was surgically implanted to allow for head restraint and a titanium recording cylinder was placed over the VLPFC as defined anatomically by (Preuss & Goldman-Rakic, 1991) and physiologically by (Romanski & Goldman-Rakic, 2002). For Subject 2, several Floating Microarrays (FMAs, Microprobes for Life Science, Gaithersburg, MD, USA) were designed specifically for the subject's prefrontal cortex and directly implanted into VLPFC during open surgery (inferior to the principal sulcus, anterior to the inferior limb of the arcuate sulcus).

Stimuli and Apparatus

In our stimulus presentation paradigm, subjects attended to naturalistic, audiovisual communication calls with their accompanying facial expressions. These vocalization movies were extracted from recordings of unfamiliar conspecifics in our home colony obtained by L.M.R. Two stimulus lists were designed such that they each

contained three identities (unfamiliar conspecifics) making three ethologically distinct expressions (Figure 1B), which differed at three levels of valance: aggressive (*pant threat*), affiliative (*coo*), and neutral (*grunt*) (Hauser & Marler, 1993; Parr & Heintz, 2009; Partan, 2002). As such, our selection of stimuli renders the type of expression (e.g. *coo*) synonymous with the valance (e.g. affiliative) and expression types are henceforth referred to by their valance. Both lists contained the same identity conspecifics, who made different exemplars of the same vocalization and facial expression. Lists alternated between recording sessions, such that the population of neurons recorded contain a mix of neural responses to one list or the other; however, the abstracted category of identity and expression were preserved in the stimuli from day to day (i.e. the second list was a copy of the first with different examples of the same identity-expression paring). The movie stimuli varied in length, based on the natural length of the vocalization and associated facial movement (average length 900ms). Additional lists of visual stimuli including objects, patterns, scenes were used during preparatory phase of a recording session.

The audiovisual movie stimuli were processed using Adobe Premiere (Adobe Systems, San Joe, CA, USA), Jasc Animation Studio (Corel, Minneapolis, MN, USA), and several custom and shareware programs. Auditory and visual components of the movies were separated into wav and mpeg streams for processing. The visual mpeg stream was edited to remove extraneous and distracting elements in the viewing frame. The audio track was also filtered to eliminate background noise if present using

MATLAB (MathWorks, Natick, MA, USA) and SIGNAL (Engineering Design, Cambridge, MA, USA). The two streams were then recombined for presentation during recordings. Stimulus audio was presented at 65-75dB SPL at the level of the subject's ear. Movies subtended 8-10° of visual angle and were presented at eye level in the center of the computer monitor.

Neurophysiology recordings were performed in a sound-attenuated room lined with Sonex (Acoustical Solutions, Richmond, VA, USA). Stimuli were presented through a video monitor flanked by speakers with ±3 dB, 75-20,000 Hz frequency response (PH5-vs, Audix, Wilsonville, OR, USA). The monitor and flanking speakers were located 29 inches from the subject's head and centered at the level of the subject's eyes. During all neural recordings, eye position was monitored using an infrared pupil monitoring system (ISCAN, Burlington, MA, USA).

Experimental Procedures and Presentation Task

Animals were acclimated to the laboratory and testing conditions. Subjects were then trained on a version of the audiovisual stimulus presentation task that did not include the stimuli used for recording (though they did encounter some stimuli in previously trained/recorded tasks). For Subject 1, the head was restrained by means of the surgically implanted post and a custom-built fixation system mounted to the primate chair. Then, either a single recording electrode (tungsten, FHC Inc., Bowdoin, ME, USA) or linear array (16ch or 32ch, V-Probe, Plexon Inc., Dallas, TX, USA) was lowered into the VLPFC each day, using a hydraulic microdrive and XY positioner (MO-95C,

Narishige Int. USA, Amityville, NY, USA) that mounted onto the recording cylinder. For Subject 2, the head was temporarily restrained with a moldable plastic headpiece while the recording headstage was connected to the FMA Omnetic connectors. Once connected, the plastic headpiece was removed and the subject was free to move through a limited range in the chair during recordings.

During the preparatory phase of the recordings and cylinder preparation, subjects viewed lists of visual objects, patterns, and scenes. For Subject 1, an electrode or linear array was lowered into the VLPFC while the subject viewed these preliminary lists. The linear array was lowered until all of the contacts were below the dural surface. Single electrodes were lowered past layer 1 until single cells could be easily isolated. For Subject 2, the implanted arrays were surgically implanted so no manual modifications could be made and 64 channels of data were recorded by default.

Wideband neural data and behavioral event codes were recorded, digitized at 40,000hz, and saved to disk using a neural data acquisition system that featured analog to digital conversion at the head stage (Omniplex, Plexon Inc.). For Subject 1, the wideband signal recorded from Plexon v-probes or Tungsten electrodes was bandpass filtered to 600-6000hz and played back through a pre-amplifier and speaker in order to optimize the position of the electrode, prior to recording. In certain instances, the 600-6000hz "spikeband" was also saved alongside the wideband data.

At the start of each trial, a red fixation point (0.15x0.15 degrees) was presented at the center of the monitor. Subjects were given a 5s period in which a 500ms sustained

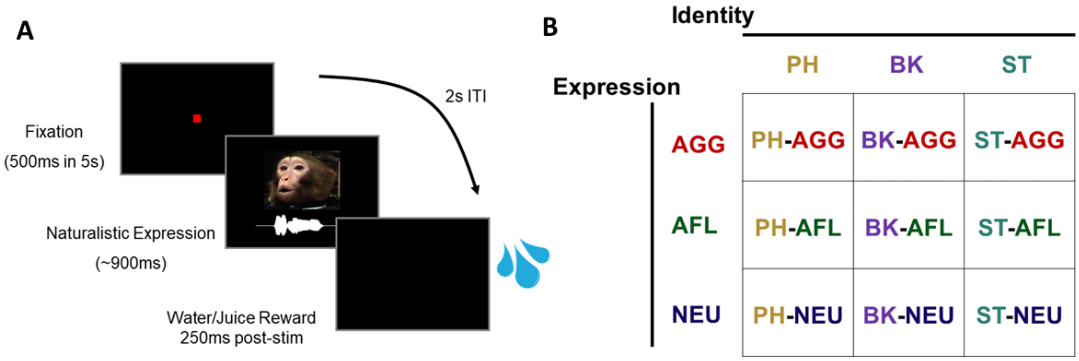

Figure 1: Naturalistic Expression Presentation Task

2 macaque subjects viewed movie clips (audio and video) of unknown conspecifics making ethologically valid expressions. (A) describes the series of events in the task. At the start of a trial, subjects were given a 5 second period in which they must fixate on a central fixation dot to trigger the stimulus. Stimuli were subsequently presented and fixation was required within the viewing window of the stimulus (10°x10°) for the length of the stimulus (900ms mean length). Stimuli varied in length due to the natural differences of timing in macaque vocal and facial expression. Reward was given 250ms after the termination of the stimulus on a successful trial, the same amount of reward was given for each stimulus, and the trial was ended at the reward. A 2 second inter-trial interval occurred before the next fixation dot. (B) The matrix shows the design of our stimulus list, color coded to match further figures. Movie clips of 3 conspecifics (PH, BK, and ST) making aggressive (AGG: pant threat), affiliative (AFL: coo), or neutral (NEU: grunt) expressions were presented pseudo-randomly. Two lists were constructed with these parameters and alternated between recording sessions, such that neurons were recorded while subjects viewed one of two exemplars of these identity-expressions pairings.

fixation on the point would trigger the beginning of the task and a vocalization movie was presented. During the movie, subjects had to maintain their gaze within the viewing window corresponding to the boundaries of the stimulus (10°x10° window) for the entire duration of the movie stimulus. Eye position was monitored continuously through the trial. Successful completion of the trial resulted in a juice or water reward 250ms after the end of the stimulus (Figure 1). Failure to establish 500ms of fixation

within the given time or failure to maintain gaze within the boundaries of the stimulus was considered a failed trial and immediately terminated the presentation of the stimulus. All trials were rewarded equally and presented pseudo-randomly for a target of 10-12 repetitions (Subject 1) or 15 repetitions (Subject 2) of each stimulus. The inter-trial interval (ITI) was 2 seconds. The timing of behavioral contingencies, presentation of stimuli, delivery of reward, and monitoring of eye position were controlled by custom software running on a dedicated computer.

Spike Sorting

Spike sorting was conducted manually, offline, after recording sessions were completed using a dedicated computer and software (Offline Sorter, Plexon Inc.). Spikes were either sorted from the wideband signal that was high-pass filtered at 600hz (4-pole Butterworth) or from saved spikeband data (see Presentation Task and Experimental Procedures). The generalized process for spike sorting started with thresholding the filtered wideband or spike band at a level where clusters of spikes were visually distinct in Principal component (PC) 1/2 , 2/3, or 1/3 spaces. Spikes were then realigned to their global minimum after the first threshold crossing. Clusters were preliminarily designated as a single unit within one of the PC spaces and then corroborated through clustering in another PC space. In rare instances, a cluster that could not be corroborated in another PC space would be found in a PC x Non-linear energy or PC x Peak-Valley space. Uncorroborated clusters had their unit designation removed and were thus not identified as single units in our analysis. Clusters that had greater than 0.2% inter-spike

interval (1ms) violations were reduced in size, typically in the corroborating space, to meet the threshold. If such a threshold could not be met, the unit designation was removed. As such, the single units analyzed in this study have, at most, 0.2% ISI and can be visually identified as clusters in at least two combinations of PC 1, 2, 3, Non-linear energy, or Peak-Valley spaces.

Identification of Responsive Units

Because our stimuli varied in the length (range 400 – 1600, average 900) due to the natural length of the associated vocalization, any period of time used identify responsive units represented a trade-off between including time after a stimulus had terminated (for shorter stimuli) and cutting off a response before a stimulus had terminated (for longer stimuli). We found units that had stimuli dependent response latencies (Figure 2) and intended to conduct sliding bin analyses on our single unit and population data. As such, we decided to look for responses over an inclusive time period and assessed responsivity to stimuli over the 0-1500ms time period, after stimulus onset. This period of time is henceforth referred to as the "stimulus period". While this was 100ms shorter than our longest stimuli, the vast majority of cells responsive at late time points were also responsive at earlier ones for shorter stimuli, which facilitated their identification using the analyses described below.

$$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \qquad (1)$$

Responsive units were identified using a sliding bin analysis to identify time bins where the Poisson probability (Equation 1) of the average spike count for a given condition was higher than the mean baseline spike count. For each unit, a mean spike count for the 200ms prior to every stimulus presentation was calculated (λ). Then a sliding bin analysis was conducted in 200ms time bins from 0-1500ms post-stimulus onset, shifting by 100ms, to derive the mean spike count in each 200ms bin for each

stimulus presentation (10-15 repetitions, spike count rounded to the nearest whole number). Equation 1 was used to calculate the Poisson probability of spiking within each bin, where λ was the mean spike count in the 200ms prior to stimulus onset across all stimuli, x was the mean spike count in a particular 200ms bin for a single stimulus of interest, and $P(x)$ was the probability of that the spike count x was would be seen within the condition and bin being analyzed given the average spontaneous spike rate λ. This calculation was repeated for all bins and for each stimulus. Units with at least one bin in one stimulus condition where the probability of spike count was less than .1 was deemed responsive in this task. This analysis was conducted using the *R* programming language. See Figure S1 for extended data and discussion why a single significant 200ms bin was chosen.

Selectivity Analysis

For each of the 285 responsive units, a selectivity index (SI) was calculated with regard to the selectivity for specific stimuli within our stimulus set, the identities in the set, and the expressions in the set. The index provides a 0-1 scale for each unit, through which the selectivity for a particular stimulus or category of stimuli can be quantified. In this analysis, a unit that selectively fired for a single audiovisual stimulus, expression, or identity would have and SI = 1 and a unit that fired uniformly to all instances would have an SI = 0. Equation 2 shows the calculation:

$$SI = \frac{n - \sum_{i=1}^{n}\left(\frac{\overline{r}_i}{\overline{r}_m}\right)}{n-1} \qquad (2)$$

where $\overline{r}_i$ represents the mean of the absolute value of the difference in firing rate between the baseline (-200ms - 0ms) and response period (0 - 1500ms) for a given stimulus, identity, or expression type i; $\overline{r}_m$ is the maximum of $\overline{r}_i$ for all i; and n is the number of stimuli (9), identities (3), or expressions types (3) for which an SI was calculated.

Two-Way ANOVA for Expression and Identity

To assess the influence of expression type and identity on the firing rates of responsive units, we fit the firing rate of each unit during the response period (0-1500ms) with a Two-Way ANOVA model of the form in equation 3:

$$r_{i,j} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j} + \sigma \qquad (3)$$

where $r_{i,j}$ is the firing rate over all expression types ($i = 1,2,3$) and identity ($j = 1,2,3$) levels, μ is the mean firing rate, α is the effect due to the type of expression, β is the effect due to identity, $(\alpha\beta)$ is the effect due to the interaction of expression type and identity, and σ is Gaussian noise. This ANOVA was performed for each responsive unit, rendering p values for the statistical significance of the effect of expression type (α), identity (β), and interaction ($\alpha\beta$) on each units firing rate. Units were thresholded as positive for each effect at the p=0.05 level for that effect. As such, units could be

significant for multiple effects. Two-Way ANOVA analysis and plotting was completed using *R* and packages *stats*, *ggplot2*, and *ggVennDiagram*.

Single Unit K Nearest Neighbor Decoding

For all 285 responsive units, a *k* nearest neighbors classifier (KNN) was used to decode the expression type or identity of a stimulus in a single trial based on the firing rate observed in that trial. For each trial that was classified, the Euclidean distance between the firing rate in the trial to be classified and all trials in a training set was calculated. The trial was then classified based on the majority classification in an odd number of nearest neighbors within the training set. Classifications were made and accuracy was calculated using 5-21 nearest neighbors. The maximum accuracy across this set of nearest neighbors was used as the accuracy for the unit being analyzed. This analysis was conducted with 3-fold cross validation (CV) with 5 repetitions within each fold averaged to produce that fold's accuracy metrics. The matrix provided to the decoder was *n* rows by 1 column + a column of labels, with *n* = number trials (typically 90 or 135 for 10 or 15 repetitions of each stimulus, respectively) and the column containing the 0-1500ms firing rate for a single unit. This analysis was implemented in *R* using additional packages *caret*, *knn*, and *tidyverse*.

Pseudo-population K Nearest Neighbor Decoding

Pseudo-populations of sizes 2-280 neurons were constructed from the responsive units. Expression type and identity were then decoded from the pseudo-population

using a KNN model, similar to that used for single unit decoding. In this case, classifications were made and accuracy was calculated using 5, 7, or 9 nearest neighbors and 2-fold CV. 50 repetitions were conducted at each pseudo-population level and each repetition included a random selection of units to include in the population and randomization of trials within the pseudo-population, in addition to standard randomization of the data for CV.

For each pseudo-population size, a matrix of 135 rows by *u* columns was created, containing the z-scored firing rates of units to 135 trials as rows (15 repetitions of 9 stimuli). In cases where 135 trials were not recorded for the unit, additional data was produced by random sampling from existing data for that unit. In the 5 cases where there were an unbalanced number of completed trials for a unit (e.g. a unit had 10 repetitions of one stimulus and 9 of another) the unit was removed from this analysis. The order of trials was shuffled prior to insertion in the matrix. *u* was the pseudo-population size and units from the 280 responsive units appropriate for this analysis were sampled at random to populate *u* columns on each iteration of the decoding analysis.

In summary, a 135 x *u* data matrix + a column of labels was the input data for the KNN decoder. Euclidean distance was used, 2-fold CV was completed on each iteration, and 50 iterations were completed for each pseudo-population size *u*. The data matrix was rebuilt, starting with random selection of units, for each iteration. This analysis was implemented in *R* using the previously mentioned packages.

Decoding of Identity and Expression Over Time

A pseudo-population of 285 was created from all responsive units by creating a 135 x 285 data matrix + a column of labels (expression type or identity) for 200ms bins (200ms bin size, 100ms step size, 16 total bins) across the response period. A single column represented the z-scored firing rate in a 200ms period of interest for a single unit in response to each of 135 trials. Random sampling of existing data was used to produce 135 trial column vectors for units that did not have 135 recorded trials (same as previous). The order of trials was shuffled for each unit prior to insertion into the matrix (same as previous). Decoding was performed on firing rates in 200ms time bins, sliding forward by 100ms, starting at 250ms before stimulus onset and ending with a bin starting at 1350ms after stimulus onset.

For each 200ms bin, the decoder generated mean vectors for each classification (either expression type or identity) from the training data and classified each trial in the test data based on the mean vector it is maximally correlated with (Pearson's correlation coefficient). This was done using 2-fold CV and repeated for 50 iterations for each 200ms bin. Mean accuracy and standard error were derived from the 50 iterations. This analysis was conducted in *MATLAB* using the *Neural Decoding Toolbox* (Meyers, 2013) and visualized in *R* using packages previously mentioned.

Principal Component, Group Distance, and Trajectory Analysis

The mean firing rate of each of the 285 responsive units was calculated for each of the 9 stimuli using the 0-1500ms response window. The rates were then z-scored and

compiled in a matrix of 9 rows x 285 columns (stimuli by units). Principal component analysis (PCA) was conducted on this matrix to project the population response to each of these 9 stimuli into 8 retained dimensions. When visualized, the coordinates of the 9 population responses were taken from their respective PC.

Distances between population responses to stimuli were calculated using the Euclidean distance between the coordinates in the preserved number of PCs. For example, the distances in a population space preserving 2 PCs were calculated for each point using $\sqrt{(PC1_{S1} - PC1_{S2})^2 + (PC2_{S1} - PC2_{S2})^2}$ where $PC1_{S1}$ is the coordinate along PC1 of the population repsonse to stimulus 1. These distances were grouped according to whether they were a distance between the same identity (ID), between the same expression type (EXP), or between mismatches of identity and expression (OTHER). A Two-Way ANOVA of the form in equation 4 was fitted to the distances calculated in spaces retaining 1-8 PCs. A post-hoc test, Tukey's Honest Square Differences (Tukey HSD) was used to specify which groups had statistically significant differences from each other.

$$d_{i,j} = \mu + \alpha_i + \beta_j + \sigma \qquad (4)$$

The trajectory of the neural population response to each stimulus was visualized by projecting the position of the population response at each time point of interest into the 8 dimension PCA space of the population at the final time point of 1500ms. First, the rotation matrix mapping the original space to the 8 PC space (**R**) was calculated from the

PCA covering the full time range fo 1500ms. For each time point of interest (i, 50ms increments from 0m-1500ms), a new population matrix ($\mathbf{P}_i$) of 9 rows x 285 columns (stimuli X units) was produced from the mean firing rate to each stimulus from 0 - i ms post stimulus onset. This matrix was rotated into the final PC space ($\mathbf{P}_i\mathbf{R}$) and visualized by retaining the respective PCs. This was repeated for all i intervals of interest to produce visualization of the path the population response to a single stimulus took over time. These analyses were conducted using *R* and additional packages *PCA*, *factominer*, and *factorextra*.

Results

Single Units Responsive to Naturalistic Stimuli Exhibit Complex and Non-Selective Responses

We recorded and isolated 466 units (single neurons and multi-units) from the VLPFC of 2 macaques, while they viewed audiovisual movie clips of 3 different conspecifics making 3 different types of expressions: either an affiliative, aggressive, or neutral expression. In the subsequent analysis, "expression" and "expression type" refers to the type of expression (*coo*/affiliative, *pant threat*/aggressive, or *grunt*/neutral) whereas "stimulus" or "stimuli" refers to the specific identity-expression combinations i.e. one or multiple of the 9 audiovisual movie clips seen and heard. Due to variability of responses to stimuli found across time (Figure 2, Figure 3), a custom filter for responsivity, using a sliding bin analysis of Poisson probability, was constructed to identify units responsive in this task (see Methods: Identification of Responsive Units). From this analysis, all units identified as responsive had at least one 200ms period where the firing rate was significantly elevated or reduced compared baseline, in response to at least one stimulus. Additionally, units that had fewer than 10 repetitions of all stimuli were removed from this analysis.

285 units met the aforementioned criteria (n=172 subject 1, n=113 subject 2). Figures 2 and 3 depict the typical activity patterns found across the population of responsive units. Units had peak responses that varied in time or had multiple peaks

across the response period and displayed varying activity patterns relative to the specific stimulus presented.

The observed dynamic firing patterns prompted a quantitative understanding of the selectivity of single units for specific stimuli, identities, and expression types. A Selectivity Index (SI) was calculated for the firing rate across 0-1500ms for each unit, with regards to selectivity for each stimulus, each identity, and each expression (see Methods: Selectivity Analysis). Using this index, each neurons response to a condition can be expressed along a 0-1 scale, with 1 signifying exclusive responsivity to a single condition and 0 signifying uniform responsivity to each condition. Neurons responsive to our task were only moderately selective for specific stimuli (SI: 0.41 ± 0.13) and less selective for identity (0.22 ± 0.13) or expression type (0.23 ± 0.13) (Figure 4). The increased selectivity for stimuli compared to identity or expression was statistically significant (Kruskal-Wallis followed by pairwise Wilcoxon rank sum, $p < 0.001$, Bonferroni corrected). This suggests that neural firing of single units in response to naturalistic expressions correlates more with combinations of stimulus features than categorical variables like identity or expression.

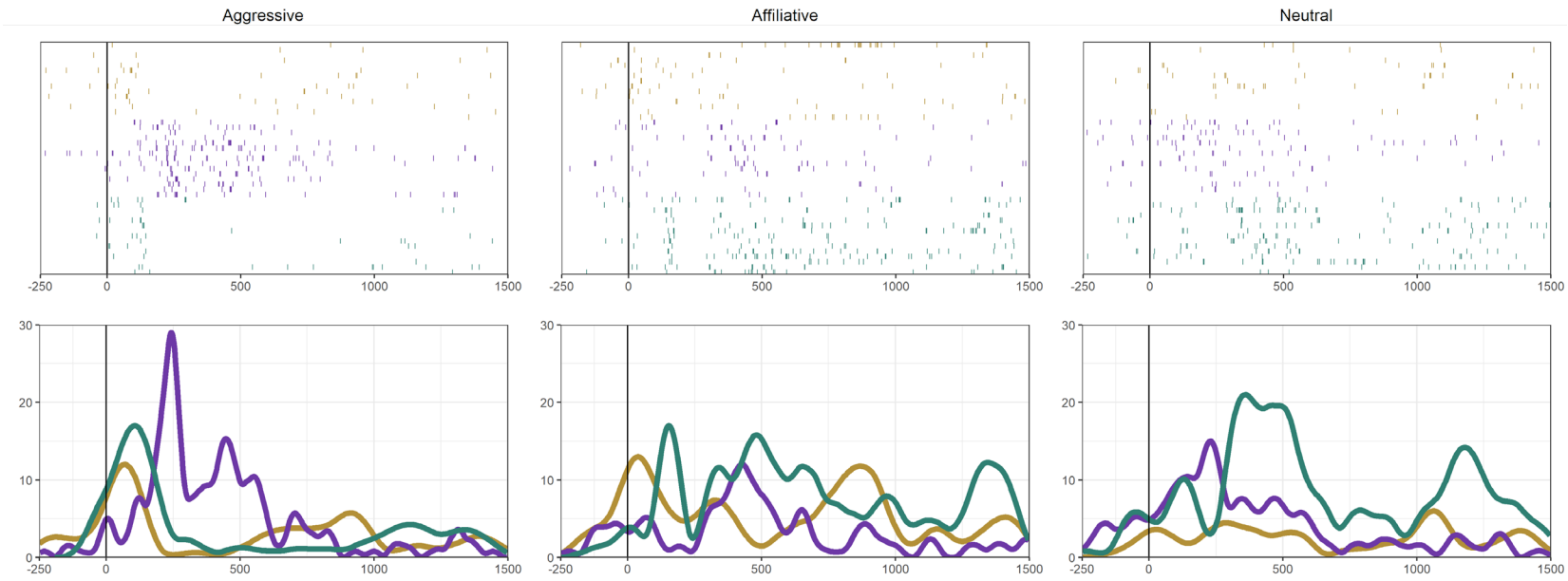

Figure 2: Single Unit Response to Multiple Presentations of Audiovisual Expressions

The responses of a representative cell multiple repetitions of all nine expressions are grouped by identity and expression. Responses to identity C1 are in gold, C2 in purple, and C3 in teal. Raster plots for each trial are presented above with corresponding SDFs below, grouped by each identity. 0ms represents the onset time of the stimuli. Because these stimuli were naturalistic expressions, consisting of video and audio component, stimuli varied from 400ms to 1000ms, based on the length of the expression. Identity was a significant modulator of the 0-1500ms firing rate for this unit, as well as the interaction between Identity and Expression (Two-Way ANOVA Identity X Expression, $p < 0.005$ for Identity, $p = 0.154$ for Expression, $p < 0.005$ for Interaction; unit: 006010821002KKSv113a)

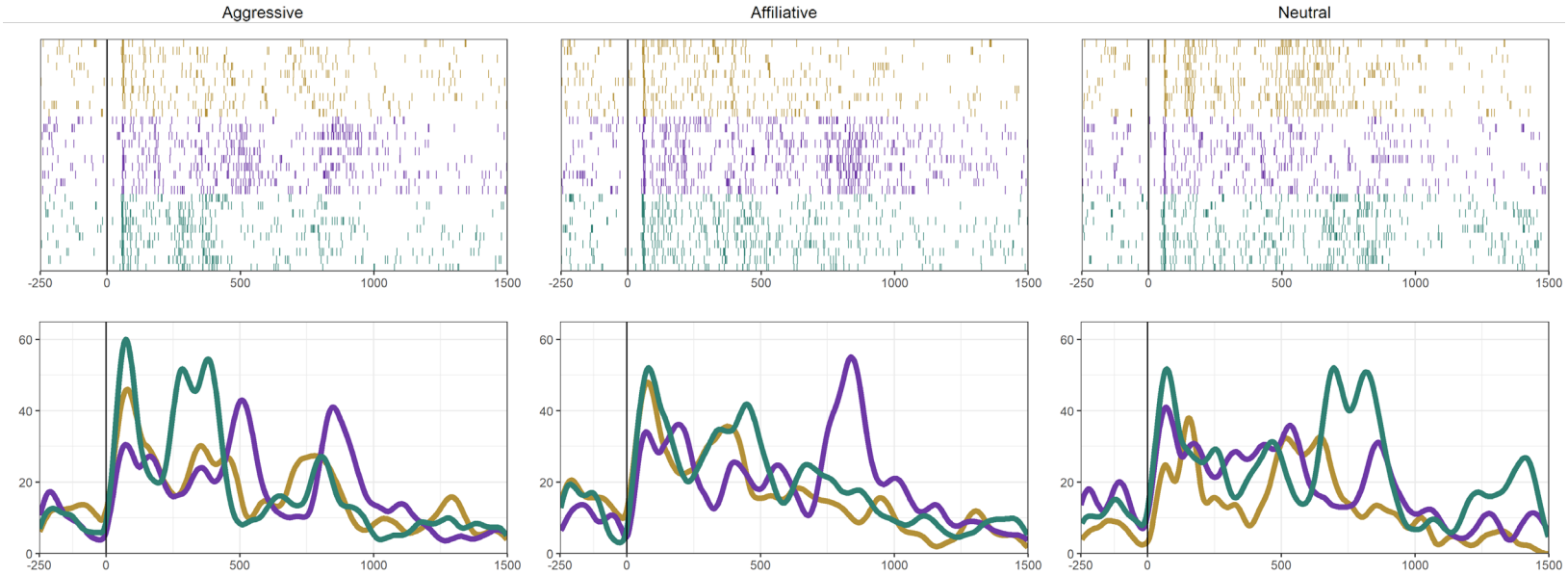

Figure 3: Single Unit with Multiple Firing Rate Influences

Another example of a responsive unit in our task. In this case, Identity, Expression, and their interaction were significantly modulated the 0-1500ms firing rate of this unit (Two-Way ANOVA Identity X Expression, $p<0.005$ for Identity, $p=0.025$ for Expression, $p<0.005$ for Interaction; unit: 320120121001ORv101a).

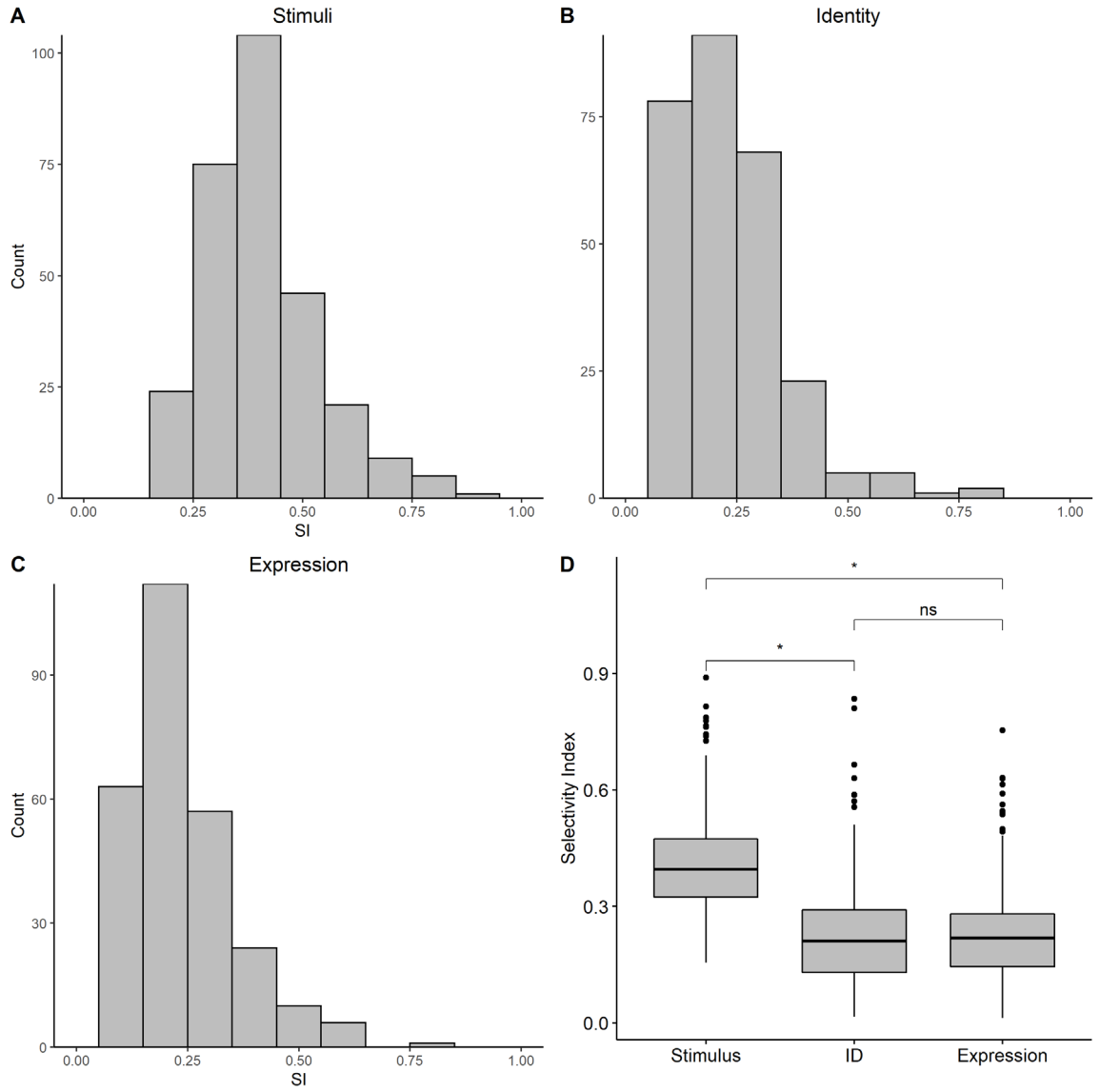

Figure 4: Selectivity Index Distributions

A selectivity index was calculated for all 285 units with regards to selectivity for one of nine stimuli, one of three identities, or one of three expressions. The plots above depict the counts of cells at each tier of selectivity index (in bins of size .1), with regard to selectivity for stimuli (A), identity (B), and expression (C). The mean of each distribution was significantly different from 0 (3 Wilcox tests, $p < 0.005$ for Stimuli, Identity, and Expression means compared to 0). Boxplots in (D) depict a comparison between means of SI for each factor of interest. The population was more selective for specific stimuli (0.41 ± 0.13) compared to identity (0.22 ± 0.13) and expression (0.23 ± 0.13) (Kruskal-Wallis, df = 2, $p < 0.001$ followed by pairwise Wilcoxon rank sum, $p < 0.001$, Bonferroni corrected). While these units are not uniformly responsive, they are also not highly selective for specific stimuli, identities, or expressions.

Despite having a higher *SI* for stimuli, the selectivity for expression type and identity across the population of units was not 0; therefore, units did not respond uniformly or indiscriminately to all identities or expressions, but they were also not selective for either variable. Thus, it was possible that these categorical variables influenced firing rates in a more subtle way. In order to further characterize the potential influence of these factors on single unit firing rates, we conducted a Two-Way ANOVA analysis modeling expression type, identity, and the interaction of the both factors as predictors of firing rate in the 0-1500ms stimulus period. A Two-Way ANOVA was conducted for each unit, and we found that 111 units from our population had a significant effect of expression type, identity, or their interaction (Figure 5A). Within the set of 111 units, 80 units had a significant effect of identity, 56 had a significant effect of expression, and 66 had a significant effect of their interaction (Figure 5B, $p < 0.05$). It should be noted that these counts represent overlapping subsets, because a single unit could have multiple significant factors in the Two-Way ANOVA. This is portrayed in the Venn Diagram in Figure 5C. The classification with the highest number of units was those for which there was a significant effect of identity, expression, and their interaction (30 units, 27% of the 111 significant units). The next highest group was those with a significant effect of identity only (28, 25%). Firing rates and *SI* for two units that showed ANOVA effects of expression type, conspecific identity, and interaction are shown in Figure 6. Overall, these results provided evidence that expression type and conspecific identity influenced the firing rates of a subset of single units.

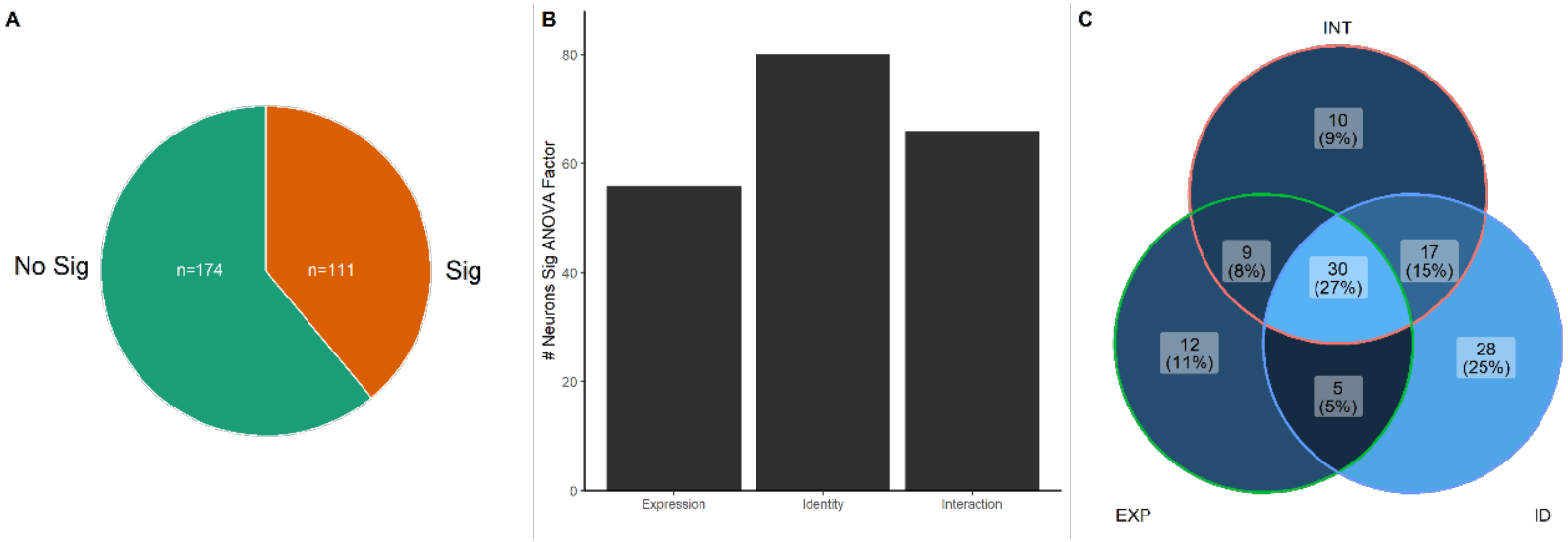

Figure 5: Two-Way ANOVA Analysis of the Effect of Expression and Identity on Single Unit Firing Rates

For each responsive unit (n=285), we conducted a Two-Way ANOVA to test for the influence of expression type (EXP), identity (ID), or the interaction (INT) of both as an influence on the firing rate of the unit. (A) Of the 285 responsive units, 111 were statistically significant ($p<0.05$) for at least one ANOVA factor (in orange), indicating an influence of EXP, ID, or INT. (B) Shows the counts of units with significant effects for each factor or interaction in this analysis. 80 units were significant for ID, 56 for EXP, and 66 for INT. (C) The units from (B) represent an overlapping population that can have multiple significant factors. Thus, there are 7 possible classifications for each unit and the distribution of these classifications is shown in the Venn Diagram. Lighter shades indicate higher counts. Percentages represent a proportion of the 111 Two-Way ANOVA positive units in (A). Overall, a co-occurrence of significant effects of ID, EXP, and INT was the highest portion in the Two-Way ANOVA positive population, indicating influences of both ID and EXP as well as specific stimuli. Examples such units with these effects are shown in Figure 6.

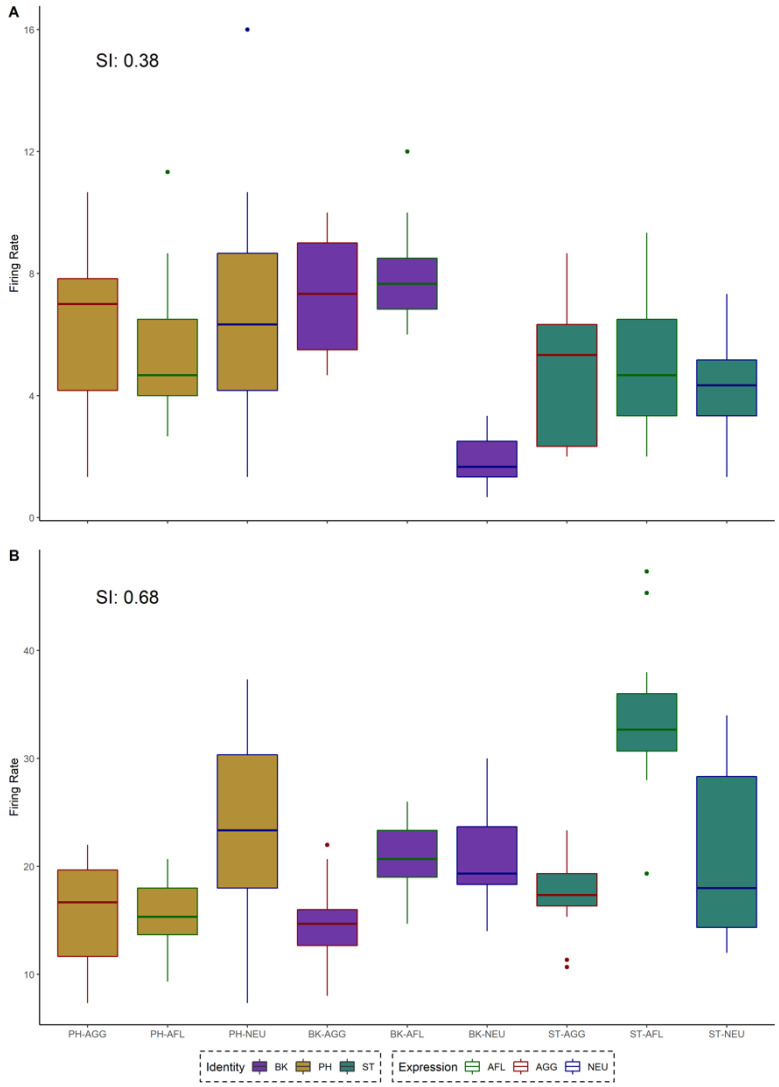

Figure 6: Firing Rate by Stimulus for Two Units with Effects of ID, EXP, and INT

Two units with box and whisker representations of firing rates for each stimulus (horizonal bar = median, box = interquartile range (IQR), stems = 1.5*IQR). Boxplot line boarders correspond to expression and boxplot fill colors correspond to identity. In both A and B, rates were calculated across 15 repetitions of each stimulus and both units had significant effects of expression, identity, and interaction in Two-Way ANOVA analysis. A Selectivity Index (SI) calculated by stimulus is provided and y axis is scaled to a relevant range for each unit. In A, a unit with firing rate trends for expressions and identity also has a single stimulus (BK-NEU) for which there is a markedly different firing rate distribution than all others. B depicts a similar trend as in A for a unit with a higher SI. In this case, stimulus ST-AFL had a markedly higher firing rate than others.

Accurate Decoding of Expression and Identity Requires a Neural Population

Since single units paradoxically exhibited both generally low selectivity and main effects of expression type and identity, we assessed whether the effects of these variables were sufficient to differentiate expressions or identities based on the firing rates of single units. We used a decoding approach in which single trials were classified by conspecific identity or expression type based on their Euclidean distance to *k* nearest neighbors (KNN model, 3-fold cross-validation (CV) with 5 repeats, maximum decoding accuracy across k = odd integers from 5-21, see Methods: Single Unit *K* Nearest Neighbor Decoding). Figure 7 depicts the results for this decoding model. The mean expression decoding accuracy averaged across the set of 285 responsive units was 0.37 ± 0.05 (significantly above chance of 0.33, $p < 0.005$, Wilcoxon Rank Sum, n = 285). The comparable mean decoding accuracy for identity was 0.38 ± 0.06 ($p < 0.005$, Wilcoxon Rank Sum, n = 285), which was also significantly higher than the mean for expression (difference of 0.01, $p < 0.005$, Wilcoxon Rank Sum). These results suggest that, on average, the firing rates of units responsive to naturalistic stimuli in the VLPFC carry some reliable information regarding expression type and conspecific identity. However, though statistically significant, the average decoding accuracy for either of these variables was only marginally above chance suggesting that single unit firing rates are not a reliable source of information. Supplemental analyses were conducted on this data to model associations between expression

and identity decoding accuracy and subject effects on decoding accuracy (see Appendix: Supplemental Information for Chapter 4).

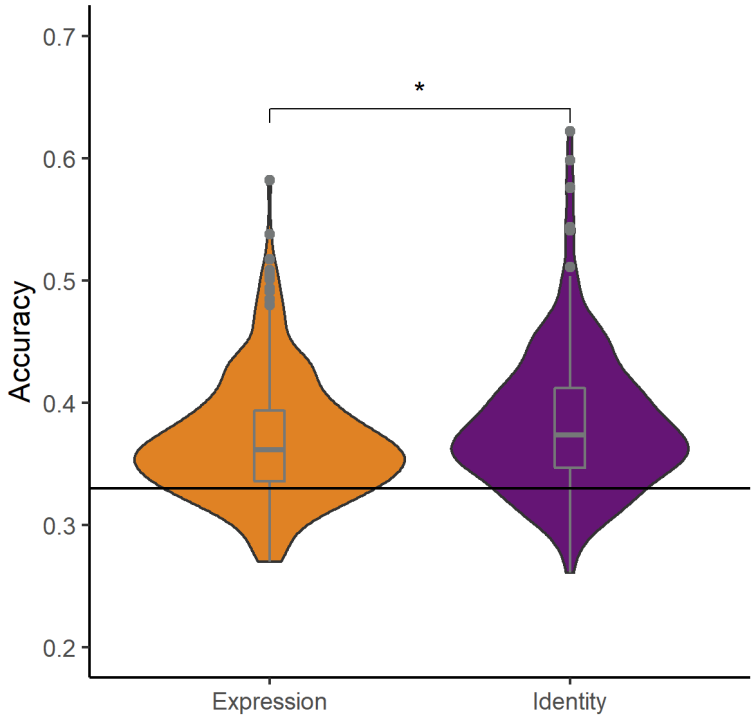

Figure 7: Weak Decoding of Identity and Expression from Single Unit Firing Rates

For each single unit, we classified single trials based on the firing rate similarity (Euclidean distance) to 5-21 nearest neighbors in training sets (KNN model, 3-fold CV, 5 repeats). The maximum decoding accuracy across 5-21 nearest neighbors was used as the single units decoding accuracy. The mean decoding accuracy of expression (0.37 ± 0.05 SD) and identity (0.38 ± 0.06) were significantly above the chance value of 0.33 ($p < 0.005$). Additionally, the mean decoding accuracy for identity was above expression (difference of 0.01, $p < 0.005$, Wilcoxon Rank Sum, Bonferroni corrected).

Although decoding of categorical social variables using single unit firing rates was largely inaccurate, the diversity of responses found in these units may facilitate accurate

decoding at the population level. To understand whether decoding of expression type or conspecific identity could be achieved by utilizing the populations of units, we conducted a decoding analysis on increasing sizes of pseudo-populations of responsive units (280 units used, 5 removed due to unbalanced number of trials). Briefly, a pseudo-population was constructed by randomly combining units to meet a target population size and randomly matching trials of the same stimuli within that population. In cases where a unit had fewer than 15 repetitions per stimulus, firing rates for that the remaining repetitions were randomly sampled from the existing data for that stimulus until 15 repetitions of each stimulus were represented in the pseudo-population for that unit. Decoding accuracy for a pseudo-population was derived from the maximum accuracy of a KNN model (see Methods: Pseudo-population *K* Nearest Neighbor Decoding) across all k (k = 3,5,7, 2-Fold CV, 50 repetitions with randomization of units in the population at each repetition). It should be noted that decoding for expression type or conspecific identity in this analysis is made more difficult by two factors inherent to the methodology of our experiment. First, single units were recorded during viewing of one of two possible sets of stimuli that contained slightly different exemplars of the same identity-expression pairings as stimuli (i.e. different exemplars of the same identity making the same expression). Thus, the decoding for conspecific identity or expression type at the population level requires a generalization of that variable across two exemplars of each stimulus. Secondly, KNN classifiers suffer from "the curse of dimensionality" (Weber et al., 1998), in which higher

dimensional spaces naturally lead to sparse representations of randomness, particularly beyond 10 dimensions.

Despite these factors, the decoding accuracy for both expression and identity increased as the population size increased (Figure 8). Expression decoding increased from .40 at a pseudo-population size of 2 to .68 at 280. Identity increased from .41 at size 2 to .80 at 280. Inspection of Figure 8 shows a non-monotonic but consistent increase in decoding accuracy as a function of population size. Additionally, the accuracy of identity was consistently above that of expression for all population sizes. This evidence supports the assertion that the decoding of expression and identity requires population level activity from single units in the VLPFC. Additionally, it suggests that information about the identity of a conspecific is abundant at the population level.

Single unit responses, like our naturalistic stimuli, were dynamic across time. Considering accurate decoding could be achieved at the population level, we sought to understand the time course of information about expression type and conspecific identity across the population. To do so, we employed a time-binned decoding model, where pseudo-population firing rates were calculated for discrete, 200ms time bins across the response period. The period from 250ms before the stimulus to 1450ms after the stimulus was analyzed in 200ms bins, overlapping by 100ms. A decoding model using the maximum correlation coefficient between a single trial and mean templates of training trials was used to classify single trials based on the 285 unit pseudo-population response (5-fold CV, 50 iterations). This model was

run for each 200ms bin in the stimulus period and for each variable of interest (expression and identity, see Methods: Decoding of Identity and Expression Across Time). The mean and 95% confidence interval of the decoding accuracy was produced and plotted for each time bin in Figure 9.

The mean decoding accuracy for identity was higher than that of expression for all time bins after stimulus onset, reaching a peak of 0.61 at the bin centered on 250ms post-stimulus onset. Expression decoding reached a peak of 0.51 at 450ms after stimulus onset. 95% confidence intervals for both variables were non-overlapping for bins centered on 250ms, 350ms, 550ms, 650ms, 750ms, and 1350ms. Overall, identity decoding increased earlier and to a higher level than that for expression. Decoding for both variables remained above chance for multiple bins after the peak of their decoding accuracy without a clear return to chance over our response period.

This suggests viewing expressive stimuli generates sustained population activity in the VLPFC that continues to carry information about conspecific identity and expression type well after a stimulus has terminated. The early peak of identity decoding followed by expression decoding is consistent with the availability of this information over time in our naturalistic stimuli.

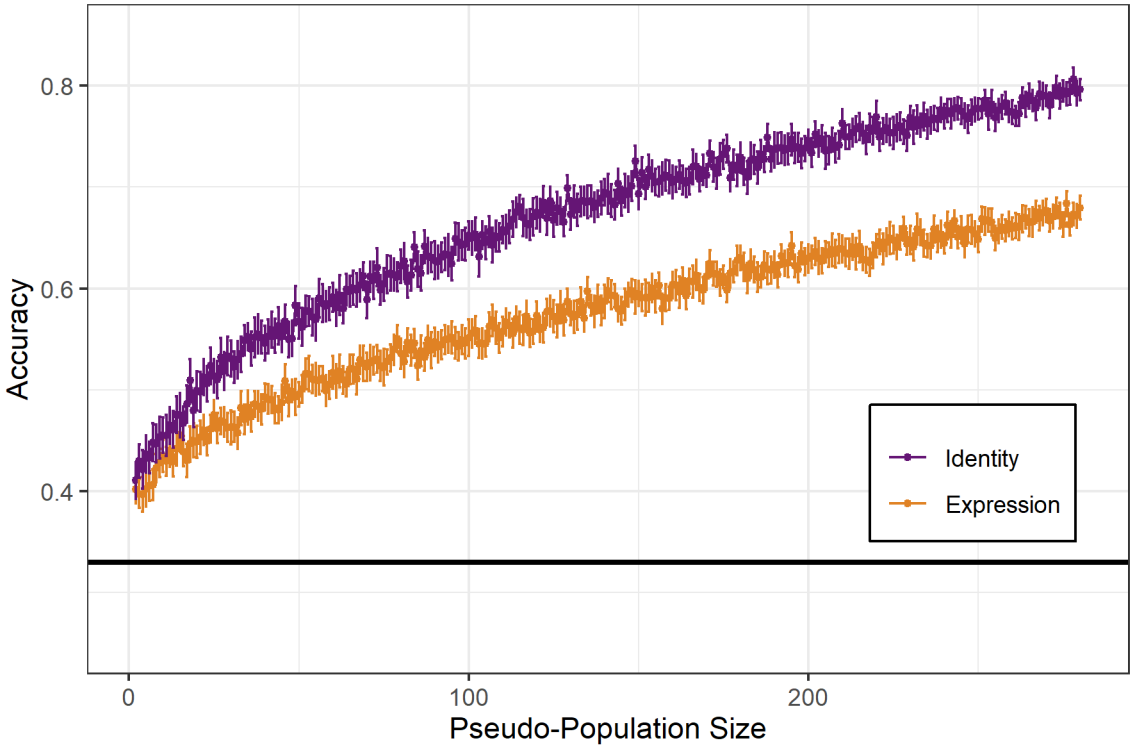

Figure 8: Pseudo-populations of Single Units Facilitate Better Decoding of Identity and Expression

Pseudo-populations of sizes 2-280 were constructed from random samples of the 280 units. A KNN decoder was used to the classify the single trials by their nearest neighbors (Euclidean distance). The maximum accuracy from classification using 5, 7, or 9 neighbors was used as the accuracy metric. At each pseudo-population size, 50 repetitions of 2-Fold CV were conducted (each repetition with a random selection of units) and averaged to produce the mean and standard error for the population size. Dots represent the mean decoding accuracy at a given population size for a particular variable, error bars represent the 95% confidence interval of the mean (± 1.96 X standard error). Decoding accuracies for both variables increase as population size increases, with identity increasing faster and ultimately having higher accuracy at larger populations than expression.

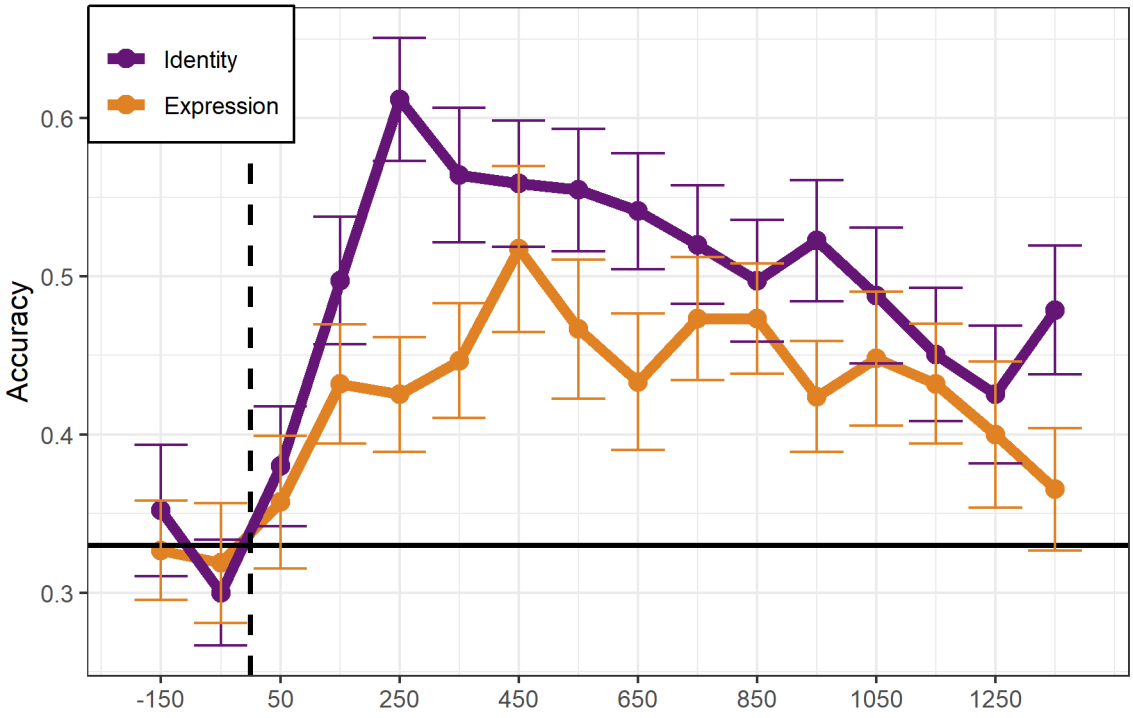

Figure 9: Sliding Bin Decoding Analysis Reveals Early Peak of Identity Decoding

A decoding analysis was performed on 285 unit pseudo-population firing rates in 200ms bins, across the response period. Dots represent the mean decoding accuracy for the variable of interest across 50 iterations of 5-fold CV. Dots represent the center of the bin (i.e. a dot at 50ms represents a bin from -50 to 150ms). Error bars represent the 95% confidence interval of the mean (1.96 * standard error). Peak decoding of identity was higher and occurred earlier than that for expression. Additionally, the mean decoding accuracy for identity was greater than expression for all bins after stimulus onset.

Identity Shapes the Population Response to Naturalistic Stimuli

Across our decoding analyses, we observed earlier and persistently robust decoding of identity across the stimulus period compared to the decoding of expression (Figures 7, 8, and 9). Given these observations, we explored the relationships of both of these categorical variables within the population activity space. In particular, we sought to understand if there was an inherent structure within this space that would facilitate accurate decoding of conspecific identity. In order to do this, we conducted a PCA on the mean firing rates of our 285 responsive units for each of the 9 stimuli (see Methods: Principal Component, Group Distance, and Trajectory Analysis). Once again, because individual units were recorded during 1 of 2 exemplars of the stimuli, this population analysis assumes a generalization of conspecific identity and expression type across two exemplars.

The variance explained by the first 3 principal components (PCs) was 47.4% (PC1: 18.8%, PC2: 15.1%, PC3: 13.5%) without a clear elbow across the 8 PCs (Figure S3). This suggests the population response space is inherently high-dimensional. Nonetheless, visualization of population responses in PC1 and PC2 displayed a trend found across all 8 retained dimensions; within PCs 1 and 2, population responses to stimuli appeared to occupy non-overlapping subspaces within the population space (Figure 10A). In other words, although a population response might be closer to that of a different conspecific identity that one of it's own (as is the case for BK-AFL and ST-AGG), responses to the same identities could be connected together

without overlapping the space occupied by other connected identities. This suggested that population responses to the same identity might be closer together than those for expression types or those for mismatching stimuli in the population space.

To quantitatively characterize this phenomenon, we measured the Euclidean distances between all population responses and grouped them according whether the distance was between population responses to the same conspecific identity (ID), the same expression type (EXP), or between responses to mismatches of both expression and identity (OTHER). We conducted this analysis for population spaces ranging from PC1 to PC8 (Figure 10B). In each population space, the median distance for the ID group was lower than the median distances for EXP or OTHER. A Two-Way ANOVA model fitting distance as a function of dimensions and group was statistically significant for both factors (dimension $p < 0.005$, group $p < 0.005$, ANOVA model: distance = dimension + group, see equation 4). Post-hoc testing with Tukey HSD revealed significant differences between ID and EXP as well as ID and OTHER ($p < 0.005$ for both). EXP and OTHER were not significantly different ($p = 0.95$). Similar trends were seen in single subject data (Figures S4 and S5).

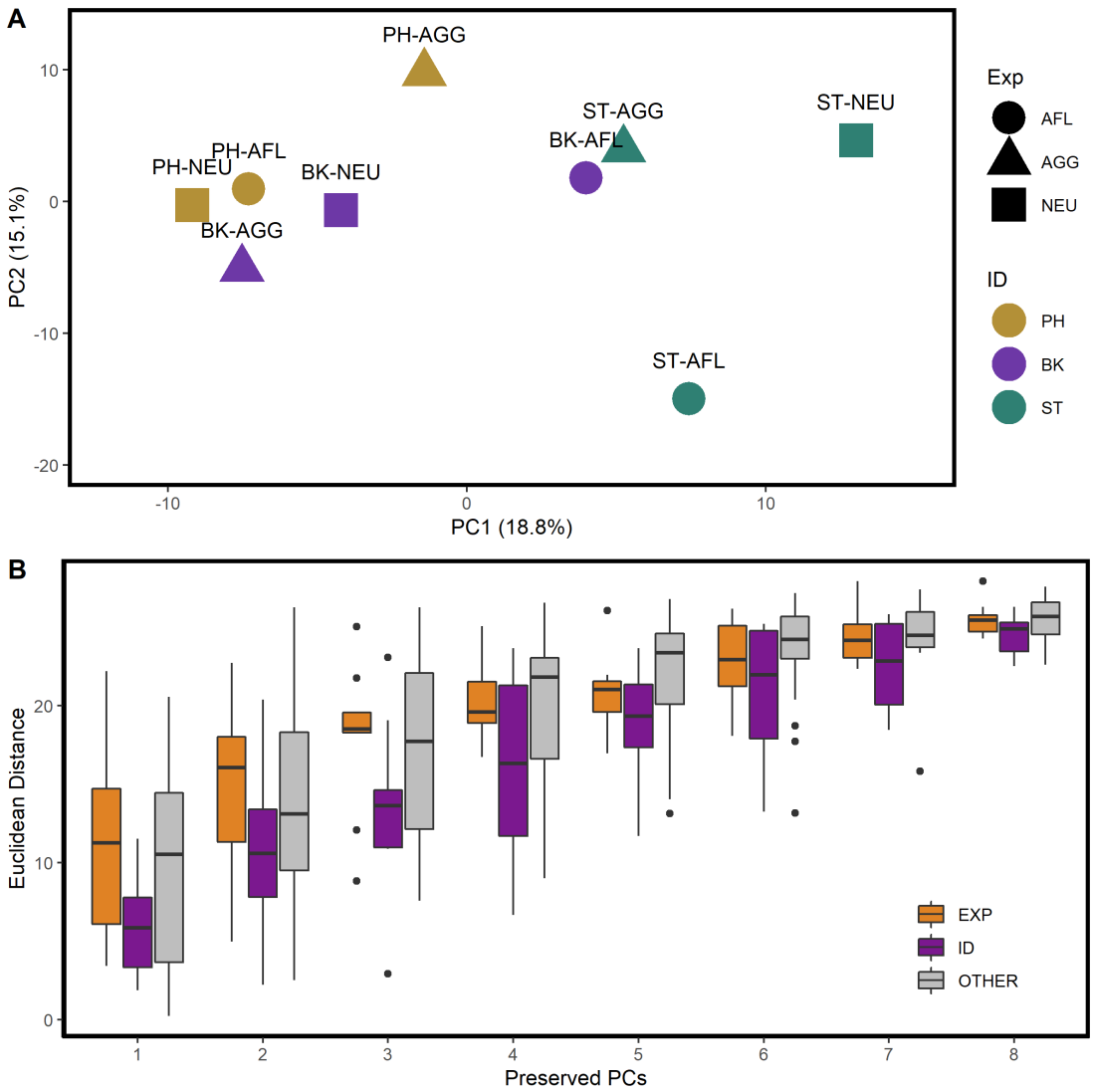

Figure 10: PCA Population Responses to Identity are Closer than Expressions

The mean firing rates over 0-1500ms for each stimuli, for all 285 responsive units, were analyzed using PCA. In A, the population responses to each of the 9 stimuli are reconstructed in PC1 and PC2. Population responses are colored by identity, to highlight that identities appear to occupy non-overlapping portions of the overall response space. In B, the distances between populations responses to similar identities (ID), expressions (EXP), or mismatching stimuli (OTHER) were measured and grouped together in reconstructed spaces preserving PCs 1-8. There were significant effects of PCs preserved ($p<0.005$) and Group ($p<0.005$) in a Two Way ANOVA (adjusted for unbalanced groups). Tukey HSD testing showed significant differences between ID-EXP ($p<0.005$) and ID-OTHER ($p<0.005$). The difference between EXP-OTHER was not significant ($p=0.95$).

Given that population responses to the same identity tended to be closer together in the population space, we pursued an understanding of the relationships between population responses across time. We particularly sought to assess whether population responses had a consistent direction of movement towards their final positions in the population response space, based on the identity of the response. To do so, we projected the mean firing rates of the 285 neuron population response at various time points along the 0-1500ms stimulus period into a common population response space (see Methods: Principal Component, Group Distance, and Trajectory Analysis). By plotting each population response's trajectory to its final position, we were able to visualize the average response of the population, over time, to each of the stimuli in our set (Figure 11). Viewed from multiple PC angles, a population response to a particular conspecific identity appears to emanate from a central region, common to all responses, and follow a trajectory, through time, that is similar to its identity counterparts. In views where responses to the same identity appear to splay in separate directions, orthogonal views corroborate that the appearance of splay is typically an artifact of the viewing angle (e.g. ST in Teal, Figure 11C shows splaying while A and B do not). This consistency of trajectory is not readily seen with populations responses to the same expression type (Figure 11D). Multiple angles of a 3-dimensional view of this population trajectory structure are provided in Figure S6.

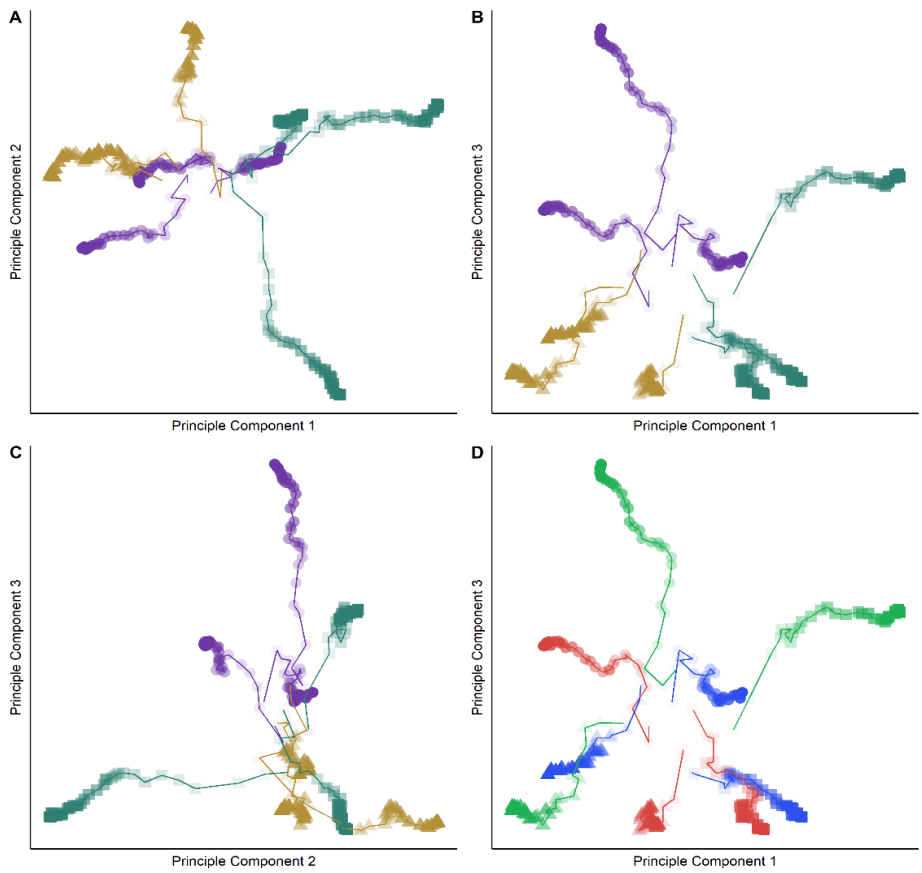

Figure 11: PCA Population Responses to Naturalistic Stimuli Follow Identity Based Trajectories

The positions of the population response to each stimulus through time were projected onto the 0-1500ms firing rate PCA space from Figure 11. Coloring of identities is the same as the previous figure, markers have been adjusted to reflect the same identity as well. Each marker represents the position of the population in at a point in time, for a given stimulus. The transparency of the marker is proportional to the time point, wherein very transparent markers represent firing rates across the early response period (e.g. 0-150ms) and less transparent markers represent firing rates across a longer response period (e.g. 0-1200ms). Trajectories are connected by lines for ease of following. (A) reflects a view of the population mean trajectories from PC1 and PC2. (B) is from PC1 and PC3. (C) is from PC2 and PC3. (D) is the same as (B), recolored to group expressions rather than identities (red = aggressive, blue = affiliative, green = neutral). The trajectories emanate from a central region and the trajectories of the same identity appear to move towards a common subspace.

Discussion

Our findings indicate that, despite a diversity of responses exhibited at the single cell level, population activity in the macaque VLPFC in response to the audiovisual expressions of conspecifics is primarily organized around the identity of the conspecific attended to. Single neurons responsive to audiovisual expression stimuli in the VLPFC exhibited complex firing patterns that varied over time (Figures 2 and 3) and exhibited weak selectivity for specific stimuli, expression type, or conspecific identity (Figure 4). Despite low selectivity, both conspecific identity and expression type exerted a systematic influence on neuronal firing rates (Figure 5) suggesting that information about both factors is transmitted by VLPFC neurons in response to the perception of an expression. Information from a neural population, rather than single neurons, was necessary to decode identity and expression type with appreciable accuracy (Figure 8). Specifically, the higher decoding accuracy of identity at all populations sizes as well as its earlier and higher peak response time, compared to expression type (Figure 9), support a model where identity is the primary factor determining the population activity of the VLPFC. Findings from our PCA analysis confirm that the structure of the aggregate activity from responsive neurons segregates population responses to the same conspecific identity into subspaces of the total response space (Figure 10) and that the activity of the population over the response period cohesively evolves such that responses to the same identity follow a similar trajectory across time (Figure 11). Together, these findings confirm the primacy of identity in

determining the population activity of the VLPFC during perception of naturalistic expressions. Furthermore, this activity is information rich, allowing for accurate decoding of expression type at the population level and potentially many more variables based on the diversity of responses seen at the single neuron level. These findings have direct implications for the study of circuits supporting normal social functions.

VLPFC Population Activity Reflects the Identity of Conspecifics

Studies of the macaque VLPFC have consistently produced evidence suggesting that the VLPFC has a critical social function. Functional neuroimaging of awake macaques viewing videos of conspecifics have shown activations in the VLPFC comparable to those in the superior temporal sulcus (STS) and other regions of the temporal lobe involved in perception of expressions (Russ & Leopold, 2015; Shepherd & Freiwald, 2018). Selectively-responsive face-cells and "face patches" have been identified in the macaque temporal lobe and VLPFC (Barat et al. 2018; O Scalaidhe 1999; O Scalaidhe et al. 1997 ; Tsao et al. 2008b), which are posited to be homologous with similar patches in humans (Tsao, Moeller, et al., 2008). VLPFC neurons are also responsive to the angle of viewed faces (Romanski & Diehl, 2011). Importantly, a factor that sets the VLPFC apart from temporal lobe face processing areas is the responsiveness to complex auditory stimuli including vocalizations and the multisensory nature of these responses. VLPFC neurons respond to species-specific vocalizations presented separately (Romanski and Goldman-Rakic 2002; Romanski et al., 2005) and combined with the accompanying facial

gestures (Diehl & Romanski, 2014; Kuraoka et al., 2015; Romanski & Hwang, 2012; Sugihara et al., 2006). VLPFC auditory cells express high selectivity for specific calls (Romanski et al., 2005) but little responsivity to pure tones and other simple auditory stimuli (Romanski et al., 2005; Romanski & Averbeck, 2009; Romanski & Goldman-Rakic, 2002).

Despite an abundance of evidence at the functional and single neuron level implicating VLPFC activity in the processing of face and vocal social stimuli, a cohesive picture of what social variables drives responses in the VLPFC still eludes our understanding. In their fMRI experiment, Shephard and Freiwald (2018) modeled functional responses along three levels; "social precepts", which contrasted viewing of scrambled images versus indirect gaze expressions; "social context", viewing indirect versus direct gaze expressions; and "facial movement", the times at which the subject produced a facial expression versus at rest (Shepherd and Freiwald, 2018). Whole brain network analysis revealed that sites in the temporal lobe were strongly associated with social percepts and sites in the medial prefrontal cortex were strongly associated with social context, while the activity of sites in the VLPFC had a moderate association to both percepts and context. Unlike the amygdala, where single units with selective responses for identities or expression have been found (Gothard et al., 2007), our study and other studies of VLPFC did not identify a population of neurons with selectivity for a particular expression type or identity (Kuraoka et al., 2015). However, such categorical representation at

the single neuron level was noted for some face cells in the orbitofrontal cortex (Barat et al., 2018).

Our combination of single unit and population level analysis illustrates the emergence of a cohesive population structure from a variety of relatively non-selective neurons. We found no neurons with a specific response for a single identity and the vast majority had lower selectivity for conspecific identity than for specific stimuli. Decoding of identity or expression by single units, on average, barely exceeded chance. However, integrated over time and number of neurons, the population activity of the VLPFC facilitated separation of different conspecific identities from complex, multimodal, time-varying stimuli. Importantly, we did not select only face-selective neurons in our recordings and analysis. Rather, we recorded from cells broadly across the VLPFC and assessed the activity of those that were responsive to naturalistic, audiovisual expression stimuli. As such, our results are, perhaps, more aligned with studies of the functional nature of the entire VLPFC under naturalistic conditions (Russ & Leopold, 2015; Shepherd & Freiwald, 2018) than those specifically targeting the face patches with a large variety of static face stimuli (Barat et al., 2018). Regardless, without selecting for sites or neurons specific to faces, and including the vocalizations that accompany prototypical facial gestures, our analysis showed that population level activity could be utilized to accurately decode both expression type and conspecific identity and that the population activity response followed a time evolving structure that aligned with identity.

The strong anatomical connectivity between the VLPFC and temporal lobe further supports our interpretation of identity as the categorical social variable that structures VLPFC population activity during viewing of naturalistic audiovisual expressions. The VLPFC is the recipient of the majority of the ventral visual stream projections to the prefrontal cortex (Barbas, 1988; Felleman & Van Essen, 1991; Kravitz et al., 2013; Petrides & Pandya, 2002) and receives auditory information in a complementary ventral stream (Romanski, 2004; Romanski, Bates, et al., 1999; Romanski, Tian, et al., 1999). Given this wealth input from the temporal lobe, it is likely that the temporal lobe and face patch system has a sizable and direct functional influence on VLPFC activity. Neurons tuned to the identity of a stimulus, invariant of the angle of gaze or orientation of the head, are typically found in the anterior face patches (Freiwald & Tsao, 2010); moreover, identity modulates the firing rate of neurons outside the traditional face patch network (Yang & Freiwald, 2021) and highly accurate decoding of identity can be achieved from small populations of cells in the anterior and middle face patches (Chang & Tsao, 2017). As such, identity is a critical social variable in the temporal face patch system that is likely equally influential for VLPFC population activity.

Expression Decoding Is Likely Facilitated by a High Dimensional Response Space

Despite the population level impact of conspecific identity and expression type, single neurons could not be clearly categorized as "identity" or "expression" responsive neurons due to inconsistent firing patterns across the exemplars of each category. This is consistent with studies

of responses to macaque vocalizations (Romanski et al., 2005), which show that single VLPFC neurons do not encode combinatorial categories across multiple exemplars. Rather, the activity patterns observed at the single neuron level resemble mixed selectivity (Rigotti et al., 2013), in which neurons have robust, typically context dependent, responses to a single or a few stimuli and are also responsive, to a lesser degree, to other stimuli in the set (Figure 6). Furthermore, VLPFC neurons exhibit non-linear multisensory integration in response to species-specific facial and vocal expressions like the audiovisual expressions used in the current study (Diehl & Romanski, 2014; Sugihara et al., 2006), which fundamentally increases the dimensionality of the population response space beyond those of the two sensory modalities (Rigotti et al., 2013). Thus, it is highly likely that the idiosyncratic response patterns seen at the single unit level are a result of mixed selectivity of sensory information that is non-linearly integrated by VLPFC. The purported utility of non-linear mixed selectivity is the simultaneous existence of a very high dimensional population response space capable of encoding a variety of unique stimuli while also providing redundant and reliable encoding that forgoes the need for neurons that are specifically responsive to a particular stimulus feature of interest (Johnston et al., 2020; Rigotti et al., 2013).

Our ability to decode expression above chance is likely the result of the existence of high dimensional space created by non-linear integration in the VLPFC. Even though we find that conspecific identity is the dominant variable structuring the population response in the first

three principal components, we were able to decode expression with increasing accuracy as more neurons were added to the population (Figure 8). Thus, while population responses to expression types seem to be diverging in lower dimensions, our decoding model (which relies on proximity in space) suggests that expression may be converging in higher dimensions.

Aside from a large expansion of the encoding space brought forth by multi-sensory integration, our use of naturalistic stimuli with synchronous, time varying auditory and visual information likely affords additional features in the time domain that are integrated into the population response space. For example, the expansion and contraction of the mouth or the abrupt auditory onset and offset of a vocalization may be features that these neurons encode. In the case of auditory information, encoding models of VLPFC neural responses to vocalizations are best constructed from Hidden Markov Models (Averbeck & Romanski, 2006), suggesting that neural firing in the VLPFC is highly dynamic across time. fMRI activation in the VLPFC during viewing of naturalistic videos (without associated audio) shows that motion is a critical correlate of activation and the impact of motion on functional activation is severely reduced when viewing non-social videos (Russ & Leopold, 2015). Most neurophysiology studies of expression and identity in the macaque employ static visual stimuli (Barat et al., 2018; Gothard et al., 2007; Tsao, Schweers, et al., 2008), present an auditory stimulus with a static image (Barat et al., 2018), and/or separate the visual movie from the vocalization (Kuraoka et al., 2015; Kuraoka & Nakamura, 2007). While doing so can afford greater control of stimulus timing and

increased variety of exemplars, it reduces the dimensionality of the population response space by failing to reproduce multisensory, time varying features inherent to the perception of naturalistic social stimuli.

Naturally, the inverse argument represents the limitations of using naturalistic stimuli. Our stimuli were constrained by the duration of the audiovisual expression and by what exemplars could be elicited from live subjects. Subtle features, such as the direction of gaze or the amplitude of the mouth movement, could not be controlled for without degrading the natural quality or timing of the expressions. Fortunately, more recent studies of macaque social perception (Khandhadia et al., 2021) are beginning to remedy this by employing 3D macaque avatars for which specific features of identity, expression, gaze, and facial action can be parametrically manipulated (Murphy & Leopold, 2019; Wilson et al., 2020). If digital social stimuli can represent the many features of macaque facial gestures which accompany their prototypical communication calls with high fidelity, it may be possible to minimize variations between stimuli and systematically test for the encoding of these features. With such methods, it is possible that responsive manipulations of stimuli based on a subjects specific population response space (akin to reassociation studies in the motor systems like Golub et al. 2018) can validate or refute the primacy of identity to the structure of population activity found in this study.

Works Cited

Averbeck, B. B., & Romanski, L. M. (2006). Probabilistic encoding of vocalizations in macaque ventral lateral prefrontal cortex. *The Journal of Neuroscience*, *26*(43), 11023–11033. https://doi.org/10.1523/JNEUROSCI.3466-06.2006

Barat, E., Wirth, S., & Duhamel, J.-R. (2018). Face cells in orbitofrontal cortex represent social categories. *Proceedings of the National Academy of Sciences*, *115*(47), E11158–E11167. https://doi.org/10.1073/pnas.1806165115

Barbas, H. (1988). Anatomic organization of basoventral and mediodorsal visual recipient prefrontal regions in the rhesus monkey. *Journal of Comparative Neurology*, *276*(3), 313–342. https://doi.org/10.1002/cne.902760302

Blair, R. J. R. (2003). Facial expressions, their communicatory functions and neuro–cognitive substrates. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *358*(1431), 561–572. https://doi.org/10.1098/rstb.2002.1220

Chang, L., & Tsao, D. Y. (2017). The Code for Facial Identity in the Primate Brain. *Cell*, *169*(6), 1013-1028.e14. https://doi.org/10.1016/j.cell.2017.05.011

Darwin, C. (1965). *The Expression of the Emotions in Man and Animals*. University of Chicago Press. http://books.google.com/books?id=IYJ9RH5PShwC&printsec=frontcover&dq=the+expression+of+emotions+in+man+and+animals&hl=&cd=3&source=gbs_api

Diehl, M. M., & Romanski, L. M. (2014). Responses of prefrontal multisensory neurons to mismatching faces and vocalizations. *The Journal of Neuroscience*, *34*(34), 11233–11243. https://doi.org/10.1523/JNEUROSCI.5168-13.2014

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129. https://doi.org/10.1037/h0030377

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1–47.

Freiwald, W. A., & Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, *330*(6005), 845–851. https://doi.org/10.1126/science.1194908

Golub, M. D., Sadtler, P. T., Oby, E. R., Quick, K. M., Ryu, S. I., Tyler-Kabara, E. C., Batista, A. P., Chase, S. M., & Yu, B. M. (2018). Learning by neural reassociation. *Nature Neuroscience*, *21*(4), 607–616. https://doi.org/10.1038/s41593-018-0095-3

Gothard, K. M., Battaglia, F. P., Erickson, C. A., Spitler, K. M., & Amaral, D. G. (2007). Neural responses to facial expression and face identity in the monkey amygdala. *Journal of Neurophysiology*, *97*(2), 1671–1683. https://doi.org/10.1152/jn.00714.2006

Hage, S. R., & Nieder, A. (2015). Audio-Vocal Interaction in Single Neurons of the Monkey Ventrolateral Prefrontal Cortex. *Journal of Neuroscience*, *35*(18), 7030–7040. https://doi.org/10.1523/JNEUROSCI.2371-14.2015

Hauser, M. D., & Marler, P. (1993). Food-Associated Calls in Rhesus Macaques (Macaca-Mulatta) .1. Socioecological Factors. *Behavioral Ecology*, *4*(3), 194–205. https://doi.org/10.1093/beheco/4.3.194

Johnston, W. J., Palmer, S. E., & Freedman, D. J. (2020). Nonlinear mixed selectivity supports reliable neural computation. *PLOS Computational Biology*, *16*(2), e1007544. https://doi.org/10.1371/journal.pcbi.1007544

Khandhadia, A. P., Murphy, A. P., Romanski, L. M., Bizley, J. K., & Leopold, D. A. (2021). Audiovisual integration in macaque face patch neurons. *Current Biology*, *31*(9), 1826-1835.e3. https://doi.org/10.1016/j.cub.2021.01.102

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26–49. https://doi.org/10.1016/j.tics.2012.10.011

Kuraoka, K., Konoike, N., Neuroscience, K. N., & 2015. (2015). Functional differences in face processing between the amygdala and ventrolateral prefrontal cortex in monkeys. *Elsevier*, *304*, 71–80. https://doi.org/10.1016/j.neuroscience.2015.07.047

Kuraoka, K., & Nakamura, K. (2007). Responses of single neurons in monkey amygdala to facial and vocal emotions. *Journal of Neurophysiology*, *97*(2), 1379–1387. https://doi.org/10.1152/jn.00464.2006

Meyers, E. (2013). The neural decoding toolbox. *Frontiers in Neuroinformatics*, *7*. https://www.frontiersin.org/article/10.3389/fninf.2013.00008

Murphy, A. P., & Leopold, D. A. (2019). A parameterized digital 3D model of the Rhesus macaque face for investigating the visual processing of social cues. *Journal of Neuroscience Methods*, *324*, 108309. https://doi.org/10.1016/j.jneumeth.2019.06.001

Niedenthal, P. M., & Brauer, M. (2012). Social Functionality of Human Emotion. *Annual Review of Psychology*, *63*(1), 259–285. https://doi.org/10.1146/annurev.psych.121208.131605

O Scalaidhe, S. P. (1999). Face-selective Neurons During Passive Viewing and Working Memory Performance of Rhesus Monkeys: Evidence for Intrinsic Specialization of Neuronal Coding. *Cerebral Cortex*, *9*(5), 459–475. https://doi.org/10.1093/cercor/9.5.459

O Scalaidhe, S. P., Wilson, F. A., & Goldman-Rakic, P. S. (1997). Areal segregation of face-processing neurons in prefrontal cortex. *Science*, *278*(5340), 1135–1138. https://doi.org/10.1126/science.278.5340.1135

Parr, L. A., & Heintz, M. (2009). Facial expression recognition in rhesus monkeys, Macaca mulatta. *Animal Behaviour*, *77*(6), 1507–1513. https://doi.org/10.1016/j.anbehav.2009.02.024

Partan, S. R. (2002). Single and multichannel signal composition: Facial expressions and vocalizations of rhesus macaques (Macaca mulatta). *Behaviour*, *139*(8), 993–1027. https://doi.org/10.1163/15685390260337877

Petrides, M., & Pandya, D. N. (2002). Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *European Journal of Neuroscience*, *16*(2), 291–310. https://doi.org/10.1046/j.1460-9568.2001.02090.x

Preuss, T. M., & Goldman-Rakic, P. S. (1991). Myelo- and cytoarchitecture of the granular frontal cortex and surrounding regions in the strepsirhine primate Galago and the anthropoid primate Macaca. *Journal of Comparative Neurology*, *310*(4), 429–474. https://doi.org/10.1002/cne.903100402

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, *497*(7451), 585–590. https://doi.org/10.1038/nature12160

Romanski, L. M. (2004). Domain specificity in the primate prefrontal cortex. *Cognitive, Affective & Behavioral Neuroscience*, *4*(4), 421–429. https://doi.org/10.3758/CABN.4.4.421

Romanski, L. M., & Averbeck, B. B. (2009). The primate cortical auditory system and neural representation of conspecific vocalizations. *Annual Review of Neuroscience*, *32*, 315–346. https://doi.org/10.1146/annurev.neuro.051508.135431

Romanski, L. M., Averbeck, B. B., & Diltz, M. (2005). Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *Journal of Neurophysiology*, *93*(2), 734–747. https://doi.org/10.1152/jn.00675.2004

Romanski, L. M., Bates, J. F., & Goldman-Rakic, P. S. (1999). Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *Journal of Comparative Neurology*, *403*(2), 141–157.

Romanski, L. M., & Diehl, M. M. (2011). Neurons responsive to face-view in the primate ventrolateral prefrontal cortex. *Neuroscience*, *189*, 223–235. https://doi.org/10.1016/j.neuroscience.2011.05.014

Romanski, L. M., & Goldman-Rakic, P. S. (2002). An auditory domain in primate prefrontal cortex. *Nature Neuroscience*, *5*(1), 15–16. https://doi.org/10.1038/nn781

Romanski, L. M., & Hwang, J. (2012). Timing of audiovisual inputs to the prefrontal cortex and multisensory integration. *Neuroscience*, *214*, 36–48. https://doi.org/10.1016/j.neuroscience.2012.03.025

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience*, *2*(12), 1131–1136. https://doi.org/10.1038/16056

Russ, B. E., & Leopold, D. A. (2015). Functional MRI mapping of dynamic visual features during natural viewing in the macaque. *NeuroImage*, *109*, 84–94. https://doi.org/10.1016/j.neuroimage.2015.01.012

Shepherd, S. V., & Freiwald, W. A. (2018). Functional Networks for Social Communication in the Macaque Monkey. *Neuron*, *99*(2), 413-420.e3. https://doi.org/10.1016/j.neuron.2018.06.027

Sugihara, T., Diltz, M. D., Averbeck, B. B., & Romanski, L. M. (2006). Integration of Auditory and Visual Communication Information in the Primate Ventrolateral Prefrontal Cortex. *The Journal of Neuroscience*, *26*(43), 11138–11147. https://doi.org/10.1523/JNEUROSCI.3550-06.2006

Taubert, J., Japee, S., Murphy, A. P., Tardiff, C. T., Koele, E. A., Kumar, S., Leopold, D. A., & Ungerleider, L. G. (2020). Parallel Processing of Facial Expression and Head Orientation in the Macaque Brain. *Journal of Neuroscience*, *40*(42), 8119–8131. https://doi.org/10.1523/JNEUROSCI.0524-20.2020

Tsao, D. Y., Moeller, S., & Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(49), 19514–19519. https://doi.org/10.1073/pnas.0809662105

Tsao, D. Y., Schweers, N., Moeller, S., & Freiwald, W. A. (2008). Patches of face-selective cortex in the macaque frontal lobe. *Nature Neuroscience*, *11*(8), 877–879. https://doi.org/10.1038/nn.2158

Weber, R., Schek, H.-J., & Blott, S. (1998). Analysis and Performance Study for Methods in High-Dimensional Spaces. *Proceedings of 24rd International Conference on Very Large Data Bases*, 12.

Wilson, V. A. D., Kade, C., Moeller, S., Treue, S., Kagan, I., & Fischer, J. (2020). Macaque Gaze Responses to the Primatar: A Virtual Macaque Head for Social Cognition Research. *Frontiers in Psychology*, *11*. https://www.frontiersin.org/article/10.3389/fpsyg.2020.01645

Yang, Z., & Freiwald, W. A. (2021). Joint encoding of facial identity, orientation, gaze, and expression in the middle dorsal face area. *Proceedings of the National Academy of Sciences*, *118*(33), e2108283118. https://doi.org/10.1073/pnas.2108283118