UNIVERSITY OF CALIFORNIA SAN DIEGO

Natural Behavior and the Neurobiology of Primate Communication

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Neurosciences with a Specialization in Computational Neurosciences

by

Vladimir Jovanovic

Committee in Charge:

Professor Cory Miller, Chair
Professor Christina Gremel
Professor Stefan Leutgeb
Professor Katerina Semendeferi
Professor Bradley Voytek

2020

The dissertation of Vladimir Jovanovic is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2020

DEDICATION

This dissertation is dedicated to:

My family: Dragan, Mira, Sandra, Aleks and Nevena. They supported me through all these years, and, in particular, to mom for patiently listening to all my ups and downs during my graduate program.

To the close friends I have made in San Diego. Even though many are moving on to bigger and better things already, I will always cherish the time we spend together.

My friends from Texas. Distance has kept us apart, but our shared experiences growing up lets us reconnect as if not a single day has gone by between us.

The Neurosciences Graduate Program that has introduced me to some of the best people I have met in my life, making this experience unique and gratifying, and worth the grueling process.

And finally, to my niece, little Katarina Jovanovic, born in the middle of writing my dissertation. I cannot wait to see you in person after I defend.

EPIGRAPH

Tamo daleko, daleko od mora,
Tamo je selo moje, tamo je Srbija.
Tamo je selo moje, tamo je Srbija.

Tamo daleko, gde cveta limun žut,
Tamo je srpskoj vojsci jedini bio put.
Tamo je srpskoj vojsci jedini bio put.

Tamo daleko, gde cveta beli krin,
Tamo su živote dali zajedno otac i sin.
Tamo su živote dali zajedno otac i sin.

Tamo gde tiha putuje Morava,
Tamo mi ikona osta, i moja krsna slava.
Tamo mi ikona osta, i moja krsna slava.

Tamo gde Timok pozdravlja Veljkov grad,
Tamo mi spališe crkvu, u kojoj venčah se mlad.
Tamo mi spališe crkvu, u kojoj venčah se mlad.

Bez otadžbine, na Krfu živeh ja,
Ali sam ponosno klic'o, Živela Srbija!
Ali sam ponosno klic'o, Živela Srbija!

Đorđe Marinković, “Tamo Daleko” 1916.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

CPP	Cocktail Party Problem
FC	Frontal Cortex/Frontal Cortical
NHP	Non-human Primate
PFC	Prefrontal Cortex
vlPFC	Ventrolateral Prefrontal Cortex
VM	Virtual Monkey

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Cory T. Miller for his support as the chair of my committee and my mentor. His constant patience, guidance, and support were invaluable.

I also want to acknowledge the help Shanna Coop gave with my monkeys for Chapter 2, and Madeleine Gagne & Victoria Ngo for the help they gave me by collecting data and retrieving wily monkeys for Chapter 3.

Chapter 1, in full, is a reprint of the material as it appears in Social Context-Dependent Activity in Marmoset Frontal Cortex Populations during Natural Conversations 2017. Nummela, Samuel U.; Jovanovic, Vladimir; Miller, Cory T.; de la Mothe, Lisa, Journal of Neuroscience, 2017. The dissertation author was one of the primary investigators and authors of this paper.

VITA

2012 Bachelor of Science, Southern Methodist University

2012 Bachelor of Arts, Southern Methodist University

2013 Master of Science, Southern Methodist University

2017-2019 Vice President of Academic Affairs, Graduate Student Association

2020 Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Nummela, S. U.,* Jovanovic, V.,* de la Mothe, L., & Miller, C. T. (2017). Social context-dependent activity in marmoset frontal cortex populations during natural conversations. Journal of Neuroscience, 37(29), 7036-7047. * Contributed equally to the manuscript

Toarmino, C. R., Jovanovic, V., & Miller, C. T. (2016). Decisions to Communicate in Primate Ecological and Social Landscapes. In Psychological Mechanisms in Animal Communication (pp. 271-284). Springer, Cham.

Vladimir Jovanovic, Margaret Dunham, Michael Hahsler, and Yu Su. “Evaluating Hurricane Intensity Prediction Techniques in Real Time.” For: International Conference in Data Mining 2011.

FIELDS OF STUDY

Major Field: Computer Science

Studies in Data Mining
Professor Margaret Dunham

Major Field: Psychology

Studies in Sensory Augmentation
Professor George Holden

Major Field: Physics

Studies in Astronomy
Professor Robert Kehoe

Major Field: Neuroscience

Studies in Neuroethology & Communication
Professor Cory T. Miller

ABSTRACT OF DISSERTATION

Natural Behavior and the Neurobiology of Primate Communication

by

Vladimir Jovanovic

Doctor of Philosophy in Neurosciences with a Specialization in Computational Neurosciences

University of California San Diego 2020

Professor Cory T. Miller, Chair

Our primate Order is known for the expansion of the neocortex relative to other mammals. This distinction is coupled with a characteristic complex society that is facilitated by dynamic social cognitive mechanisms and systems of communication. Because of this intricate relationship, investigating the neural basis of communication within primates affords the opportunity to better understand how different dimensions of sociality are supported by the

structures of the brain itself. Much of the research on the neuroscience of communication in primates has hinged on studies of vocalization processing in head-restrained monkeys either passively listening to stimuli or engaged in a conditional behavioral task. But the information communicated by social signals are heavily influenced by the natural contexts they occur in, and auditory processing of vocalizations within the brain may likewise be heavily affected by the context in which conspecific vocalizations are heard; thus, the experiments may not fully capture the neural basis of communication. I hypothesize that the traditional experimental contexts typical of nonhuman primate neuroscience research has divorced the signal from its natural context, and, consequently, limited our understanding of how various neocortical structures support these processes. Here I sought to address this critical gap in our knowledge by implementing novel experimental paradigms designed to explicate the neurobiology and behavior of natural communication in freely-moving marmoset monkeys (*Callithrix jacchus*). In this dissertation, I detail the results of new insights gained from the innovative experiments that support my hypothesis. Chapter 1 shows how broad 'states' of neural populations in frontal cortex during natural, untrained behavior of antiphonal conversations in the marmoset predicts whether subjects respond to a conspecific call. Chapter 2 shows robust within-neuron differences in how prefrontal cortex neurons respond to vocalizations between traditional head-restrained contexts and natural behavior suggesting that data recorded in the former context is not predictive of the latter. Finally, Chapter 3 shows my novel multi-speaker paradigm that simulates the natural communication networks in marmosets (i.e. "Cocktail Party") to study the vocal processing of marmosets in complex acoustic environments previously inaccessible to researchers for any other animal model. Results demonstrate that marmosets employ similar perceptual mechanisms as humans to communicate in these dynamic acoustic and social

landscapes. These findings establish a novel paradigm in which to explore the neurobiology of primate communication in dynamic, multi-speaker communication networks that more closely resemble their natural communication systems.

1 Social context-dependent activity in marmoset frontal cortex populations during natural conversations

1.1 Abstract

Communication is an inherently interactive process that weaves together the fabric of both human and nonhuman primate societies. To investigate the properties of the primate brain during active social signaling, we recorded the responses of frontal cortex neurons as freely moving marmosets engaged in conversational exchanges with a visually occluded virtual marmoset. We found that small changes in firing rate (~1 Hz) occurred across a broadly distributed population of frontal cortex neurons when marmosets heard a conspecific vocalization, and that these changes corresponded to subjects' likelihood of producing or withholding a vocal reply. Although the contributions of individual neurons were relatively small, large populations of neurons were able to clearly distinguish between these social contexts. Most significantly, this social context-dependent change in firing rate was evident even before subjects heard the vocalization, indicating that the probability of a conversational exchange was determined by the state of the frontal cortex at the time a vocalization was heard, and not by a decision driven by acoustic characteristics of the vocalization. We found that changes in neural activity scaled with the length of the conversation, with greater changes in firing rate evident for longer conversations. These data reveal specific and important facets of this neural activity that constrain its possible roles in active social signaling, and we hypothesize that the close coupling between frontal cortex activity and this natural, active primate social-signaling behavior facilitates social-monitoring mechanisms critical to conversational exchanges.

1.2 Introduction

Social factors are thought to have had a considerable impact on the evolution of the primate brain (Dunbar, 2003; Miller et al., 2016; Platt et al., 2016). Unique circuits for social signal processing and cognition, such as faces and language (Hickok and Poeppel, 2004; Tsao et al., 2006; Hung et al., 2015), reflect the potential significance of sociality in shaping many aspects of primate brain architecture. Yet, despite evidence of remarkably complex social behaviors in nonhuman primates that likely rely on this intricate neural circuitry (Cheney and Seyfarth, 2007; Rosati et al., 2010), notably few neurobiological studies directly link neuronal processes to these characteristic natural behaviors. Neuroimaging and neurophysiological studies of social communication in primates have typically presented restrained subjects with static social stimuli (e.g., faces, vocalizations, etc.; Leopold et al., 2006; Perrodin et al., 2011; Fisher and Freiwald, 2015). Because of the intrinsic interactive nature of communication, this approach effectively divorces the signal from the very social interactions they evolved to mediate, thereby limiting interpretations of these data to facets of signal processing. Not only does the social context in which social signals are produced have a profound influence on what is communicated (Engh et al., 2006; Seyfarth and Cheney, 2014), but active communication is known to affect properties of neural activity (Stephens et al., 2010; Hasson et al., 2012; Silbert et al., 2014). Because of the sophistication of the primate social landscape, and the evolution of neural circuits to support these behaviors, neurobiological studies of active communication are likely to yield unique insight into the neural processes supporting distinct aspects of the primate brain related to social functions (Hasson et al., 2012; Miller et al., 2016).

Primate communication might be based not only on the content of individual social signals, which are limited in number and content, but also on communicative behaviors that mediate myriad social interactions characteristic of their societies (Miller et al., 2016). Marmoset antiphonal conversations, a naturally occurring vocal behavior characterized by the coordinated reciprocal exchange of phee calls (Fig. 1; Miller and Wang, 2006; Roy et al., 2011), offer unique opportunities to investigate these more social dimensions of primate communication at a neurobiological level (Eliades and Miller, 2017). For example, two recent neurophysiology experiments showed that neurons in multiple areas of marmoset prefrontal and premotor cortices exhibited little to no response to hearing phee calls during antiphonal conversations, despite the same population showing robust vocal motor-related changes in activity (Miller et al., 2015; Roy et al., 2016). Notably, these findings contrasted with prior neurophysiology studies of head-restrained rhesus and squirrel monkeys showing strong sensory-driven responses to vocalizations in the same areas of the frontal cortex (Newman and Lindsley, 1976; Gifford et al., 2005; Romanski et al., 2005). The disparity evident in these findings is difficult to currently reconcile, but suggests that, like human communication (Hasson et al., 2012), natural primate communication may involve processes that are not strictly sensory and motor.

Further analyses revealed a potentially distinct, parallel mechanism to sensory encoding in the marmoset frontal cortex during active communication. We found that frontal cortical activity when subjects heard a phee call could classify whether subjects produced a subsequent response or not in the conversation, despite the dearth of stimulus-driven activity evident at the level of single neurons (Miller et al., 2015). This intriguing result suggests that the frontal cortex participates in the outcome of marmoset conversations, but a more thorough characterization is required to distinguish among the many mechanisms at play during active vocal interactions.

These mechanisms include sensory encoding, perceptual categorization, decision making, attention, and arousal. Here we thoroughly characterize the underlying sources of variance in frontal cortical activity, narrowing its possible role in natural conversations. By doing so, we take important steps toward understanding a specific neural mechanism in the technically and conceptually challenging context of natural, freely moving, primate social behaviors.

1.3 Methods

1.3.1 Subjects

Three adult common marmosets (Callithrix jacchus) group-housed in the Cortical Systems and Behavior Laboratory at University of California, San Diego served as subjects in these experiments. Marmosets are a New World monkey endemic to the forests of northeastern Brazil (Schiel and Souto, 2017). Marmoset Subjects B and R were male. Marmoset Subject F was female. We recorded neural activity from two microelectrode arrays in Subject B. The array in the left hemisphere, B01, was centered in area 6v, while the second array, B02, was centered in area 6d in the right hemisphere. Subject R had a single array, R01, placed in the right hemisphere centered in areas 45 and 8av. Subject F had a single array, F01, placed in the left hemisphere centered in area 6d with the most rostral electrodes in 8ad, similar to array B02. Microelectrode array locations were chosen based on previous functional neuroanatomy study of marmosets engaged in natural vocal communication (Miller et al., 2010b).

1.3.2 Surgical Procedures

Before the placement of the electrode arrays and initiation of the neurophysiology experiments, all subjects underwent a surgery to implant an acrylic head cap and stainless-steel head posts. During this surgery, the lateral sulcus, as well as the rostral and lateral edges of frontal cortex, were visible through the skull and marked. We were able to later use the markings

on the skull made during surgery to triangulate the desired location of the frontal cortex when placing the microelectrode array. We recorded neural activity using a Warp16 electrode array (Neuralynx). The Warp16 comprises 16 independent guide tubes that house sharp tungsten electrodes (impedance, 2.5–3.5 MΩ) in a 4 × 4 mm grid. Since the arrays are positioned on the surface of the brain, electrodes are lowered perpendicular to the laminar surface of the neocortex. Individual electrodes in the Warp16 were advanced incrementally over the course of the experiment by restraining animals in a monkey chair. A calibrated Warp Drive pusher was attached to the end of each guide tube and each respective electrode was advanced 10–20 μm twice a week. The Warp16 array was coupled with a tether to allow for freely moving behavior during recordings.

1.3.3 Experimental Design and Statistical Analysis

1.3.3.1 Behavioral Paradigm

All recordings took place in a 4 × 3 m radio frequency-shielded testing room (ETS-Lindgren). A speaker (Polk Audio, TSi100; frequency range, 40–22,000 Hz) was placed 5 m away on the opposite side of the room with cloth occluders equidistant between the animal and speaker. All vocal signal stimuli were broadcast at 80–90 db SPL measured 1 m in front of the speaker. A directional microphone (Sennheiser, model ME-66) was placed 0.5 m in front of the subject to record all vocalizations produced during a test session. For each behavioral session, marmosets were removed from colony housing <1 h before the session, and returned to the colony after the session was complete between 9:00 A.M. and 4:00 P.M. (the colony had a 6:00 A.M. to 6:00 P.M. light cycle), with each subject run at the same time of day. Further details of the playback and software are provided in previous publications (Miller and Wang, 2006; Miller

et al., 2009, 2015; Miller and Thomas, 2012). Here we briefly describe the overall procedure used during these experiments.

Marmosets produce phee calls both within antiphonal conversation and independent of these vocal interactions. Based on previous behavioral studies (Miller and Wang, 2006; Miller et al., 2009; Chow et al., 2015), phee calls that receive a marmoset response within 1–10 s of hearing it are deemed antiphonal, while calls that do not elicit a timely response are classified as independent (Fig. 1). Thus, the social context (antiphonal or independent) of a phee stimulus is determined by events after the call has been heard; that is, by whether the subject vocally responds. Importantly, there is no evidence that the acoustics of the phee call determines its social context, as the use of a discriminant function analysis was unable to distinguish between phee calls produced in these two contexts (Miller et al., 2010a). Our primary interest was comparing the impact of the two social contexts of the phee stimuli on frontal cortical activity.

In each recording session, stimuli were phee calls produced by a single marmoset previously recorded during naturally occurring antiphonal calling interactions. Our interactive playback software was designed to broadcast these stimulus classes, antiphonal and independent, at different intervals relative to subjects' behavior. Each time a subject produced a phee call, an antiphonal phee-call stimulus was broadcast 2–4 s following call offset. Bouts of antiphonal calling occurred when subjects alternated an antiphonal call response with a stimulus presentation successively, which we refer to as an extended conversation. Independent phee-call stimuli were broadcast if subjects produced no phee calls for 45–60 s. The aim of broadcasting independent stimuli was to induce conversational exchanges in subjects. Only phee calls with two pulses were analyzed. All stimuli produced by the virtual monkey consisted of two pulses, and one-pulse and three-pulse calls by subjects were extremely rare (<1% of data).

1.3.3.2 Spike Extraction and Sorting

Neural activity was digitized and sorted off-line. Based on previous reports using similar recording methods (Eliades and Wang, 2008a,b), units were determined based on the criteria that the unit have a signal-to-noise ratio (SNR) ≥13 dB and, after spike sorting, that the waveforms appeared throughout an entire recording session, which typically lasted 60–80 min. Units with <1% of interspike intervals within a 1 ms refractory period were classified as single units, and all others were classified as multiunits. Multiunits typically occurred when spike sorting was unable to separate several lower-amplitude waveforms. We used the activity of all single and multiunit recordings from sessions with ≥20 independent and antiphonal stimuli.

1.3.3.3 Simulations of Single and Population Recordings

Simulations of individual and population responses were performed for further analyses, including principal components analysis (PCA), and two-means classification. For individual units, we performed nonparametric Monte Carlo simulations of the firing rates in response to phee calls by drawing responses to 5000 stimuli, with replacement, evenly divided between antiphonal and independent stimuli. Firing rates were calculated during four time periods, each close to 1.5 s long, relative to each stimulus (Pre: 1.5 s before stimulus onset; Voc 1: first stimulus pulse; Voc 2: second stimulus pulse; Post: 1.5 s immediately following stimulus offset). We calculated the z score of firing rates for both independent and antiphonal stimuli for each time period from each unit so that all dimensions were centered for further analyses. Firing rate draws were always conserved across time periods (i.e., firing rates for Pre and Voc 1 time periods were always from the same phee stimulus). Although Monte Carlo simulation for individual units was unnecessary, it preserved any influence the process may have had on population simulations when comparing two-means classification. For simulating population

responses, one response (firing rates over all four time periods) was randomly drawn from the same phee context from each unit. This was repeated 5000 times, with replacement, evenly split between independent and antiphonal stimuli. Thus, each population response could include responses from many different stimuli, so long as the vocalization context was the same, which was necessary because individual behavioral sessions typically included simultaneous recording of <10 units. The use of 5000 Monte Carlo samples was validated by examining the variance in two-means classification and receiver operating characteristic (ROC) analysis, increasing sample size until variance plateaued (which had occurred by 3000 samples).

1.3.3.4 PCA

Principal components and their coefficients for recording simulations were obtained using the Matlab (Mathworks) "pca," using the singular value decomposition method.

1.3.3.5 ROC Analysis

ROC analysis was applied to test simulations in principal components of the training simulations by sliding a criterion from the lowest to greatest response value in 1/1000 increments of the range, with responses greater than criterion categorized antiphonal and those less than criterion as independent, with this axis flipped if the median independent response from the training set was greater than antiphonal. Hits were correctly identified antiphonal responses and false alarms were independent responses identified as antiphonal, and the ability to separate contexts was measured from the area under the resulting curve of hits against false alarms. We repeated the entire procedure 500 times to produce confidence intervals (CIs) via Monte Carlo cross-validation. This cross-validation method, which is closely related to the bootstrap and jackknife, is more clearly applicable for this case of combining responses across multiple behavior sessions.

1.3.3.6 Two-Means Classification

Because the principal component (PC) 1 of population simulations showed such clear separation between antiphonal and independent phee calls, we devised a way to test how well we could classify the social context from PC1 of population and individual unit response simulations. We first split the firing rates to antiphonal and independent stimuli into two sets: a training set (50% of the data) and a test set (50%). This was done before the simulation of the recordings to preserve independence of the datasets. PC1 was extracted from the training dataset and two-means clustering was performed using the "kmeans" Matlab function, which determined the direction of antiphonal and independent calls. The test dataset was transformed into PC1 of the training set and two-means clustering was performed on the transformed test values. The identity of each cluster from the test dataset was assigned based on the training-set clusters (e.g., if the lower-valued training-set cluster corresponded to independent phee calls, then the lower-valued test-set cluster was assumed to also be independent phee calls). Accuracy was calculated by taking the sum of correctly identified contexts divided by the 5000 total responses in the test set. CIs were estimated by repeating 500 population simulation cross-validations. Variance in classifier performance was identified according to how the training and test datasets were split. We found 200 cross-validations were sufficient to estimate median accuracy and 95% CIs (<1% changes in estimates).

The same two-means classification was also used on individual units and individual sessions using the exact same procedures, except the dimensionality of the data was reduced by including fewer units. For sessions, Monte Carlo population response simulations were performed with (normal) and without (shuffled) drawing responses for each unit from the same

stimulus (and not just within the same context). When combining units across sessions, responses must be drawn from different stimuli (although still within the same context).

1.3.3.7 Determining Stimulus Preference for Individual Units

PC1 coefficients from population training simulations were used to define the preferred stimulus of each recording. This method was reliable in that the axis of PC1 was preserved across training and test datasets for all 500 simulations. To do this, preference was assigned based on the sum of PC1 coefficients over all time periods. For most of the training datasets (98%), antiphonal preference was assigned to positive values and independent to negative values. Importantly, all analyses that involved calculating a score from responses, or that involved combining responses, based on unit preference only included the half of the stimulus set presented to each unit that was not used to calculate the stimulus preference. This reduces the number of trials available for the analyses, but it is necessary to prevent the stimuli used to calculate preference from biasing subsequent analyses in favor of that preference. Z score was used to normalize all unit responses. Significance of context preference index for individual units used the distribution of indices for each unit from the 500 Monte Carlo cross-validations, applying a one-tailed criterion with $\alpha < 0.05$, for indices grater than 0. For comparing preferential activity across populations of units, we performed t tests on the median normalized firing rates of all Monte Carlo cross-validations, which had unimodal central tendencies, with degrees of freedom determined by the population of 258 units.

1.3.3.8 Measurement of Neuronal Correlations

To estimate the correlation in activity between units, we looked at each unit, with at least one other simultaneously recorded unit (n = 256 units, because two behavioral sessions included only one unit). Pairwise correlation coefficients were calculated between each unit and all the

other units in that session, comparing firing rates for each time period (Pre, Voc 1, Voc 2, and Post) of each stimulus. The average pairwise correlation for each unit was estimated by the mean absolute value of all its pairwise correlation coefficients.

1.3.3.9 Conversation Categorization

Context preference of each unit was estimated using half the stimulus responses from each context. The other half was processed and tagged with independent and antiphonal bouts (bouts referring to consecutive stimuli of the same context). Each sequence was counted to determine bout length. A bout-related response for each unit was calculated by averaging firing rates for each stimulus over all time periods, normalizing firing rates by taking z scores across stimuli from both contexts, and rectifying responses by inverting these responses for units with antiphonal context preferences. Bout-related responses took the mean response over all stimuli that met the following bout criteria: the first and last stimuli in antiphonal and independent bouts; the second and second-to-last stimuli in independent bouts; the third and third-to-last stimuli in independent bouts (all n = 258); and, in antiphonal bouts, “middle” stimuli that were not the first or last stimulus (n = 220). Population responses and CIs were calculated from the mean and t distributions from all unit responses.

Repeated-measures two-way ANOVA was used to determine significance across six time points in bouts with factors of array location and bout category. The six time points were first and last in a bout, second and second from last, and third and third from last. Post hoc multiple comparisons with Tukey–Kramer correction were used to determine which of the bout positions within categories was significantly different from the others.

Bouts of various lengths were compared to see how population responses, as calculated above, changed depending on bout length. In each unit, only bouts of length 2–9 were analyzed, and only units with data for both antiphonal and independent bouts of the same length were included for these comparisons. Only independent bouts occurred in sequences >9, so those stimuli were not included.

Due to the decreasing sample size of the number of units for higher bout lengths, the distributions became less normal and had increasing variance. Multivariate ANOVA and ANOVA were not suitable for this. Rather, significant-difference testing was done with multiple paired-sample one-tailed t tests, which were then corrected for multiple comparisons by the Holm–Bonferroni method. Our alternative hypothesis was that mean independent bouts would be greater than mean antiphonal bouts due to the rectification of unit responses based on context preference.

1.4 Results

Our primary interest in the current study was to understand, by examining the frontal cortex population responses from three marmoset subjects, the source of variance that made it possible to predict the social context of a phee stimulus (Miller et al., 2015). One hypothesis posits that changes in frontal activity may be stimulus driven, reflecting decisions in response to hearing and encoding the phee call. Alternatively, the observed change in neural activity may also reflect a change in state unrelated to the phee stimulus. Such changes in activity could depend on many neurons distributed broadly across frontal cortical areas or a smaller proportion of neurons confined to one area. As a first step, we performed PCA on combined responses of all units to antiphonal and independent phee stimuli (see Materials and Methods, Simulations of single and population recordings). Figure 1.2 demonstrates that PCA identified a structure in the

frontal population activity that was able to separate antiphonal from independent stimuli. Figure 1.2A (top) shows a sample test simulation of frontal cortex population responses to phee stimuli plotted in PC1 and PC2 of the training simulation. Notably, the two social contexts form two clusters in PC1. As a negative control, we performed the same analysis, except that the antiphonal and independent designations for each stimulus were randomly shuffled. As expected, PCA did not separate frontal population responses by these arbitrary phee contexts (Fig. 2A, middle). To discover whether frontal cortex population responses might also distinguish between basic acoustic features of phee calls, we performed the same PCA analysis, except that stimuli were categorized by phee stimulus length instead of social context (Fig. 2A, bottom). As with the arbitrarily assigned contexts, PCA did not separate population responses by stimulus length.

We used a ROC analysis to measure how well each PC of a training simulation separated population responses of the test simulation (see Materials and Methods, ROC analysis). An area under the ROC of 0.5 indicates no separation of responses and an area of 1 indicates perfect separation. Figure 1.2B plots the median area under ROC for population responses to social contexts (top), to the randomly assigned contexts (middle), and to phee stimuli by length (bottom) for the first three PCs. PCs 1 and 2 separated population responses to antiphonal stimuli from independent stimuli to a significant degree (Monte Carlo cross-validation, $p < 0.002$, the minimum p value definable given 500 cross-validations), with greater separation in PC1 (median, 0.96) compared with PC2 (median, 0.75; Monte Carlo cross-validation, $p < 0.002$). No individual PC (or combination of PCs) significantly separated population responses to randomly shuffled contexts or by phee-stimulus length.

To better understand how the population activity was able to distinguish between antiphonal and independent contexts, we examined the coefficients assigned to each dimension of the population responses. Figure 1.2C shows the median PC1 coefficients from 500 training simulations, organized by unit in columns and by time period in rows. We presented PC1 median coefficients because they were unimodal, with a strong central tendency over the simulations resulting in highly significant correlations between PC1 coefficients between simulations (mean of $r(1030) = 0.59$, all p's < 0.0001). Units were sorted in order of mean coefficient magnitude across all four time periods, and half of the coefficient contributions were from the 78 most strongly weighted units, implying that many units contribute to the distinction between phee contexts. Notably, coefficients are exceptionally evenly distributed over time periods (Fig. 2D), so that when averaged over all units, no single time period showed a greater contribution to PC1 than any other [$t(257) < 0.88$ (magnitude), $p > 0.38$, no correction for multiple comparisons]. This suggests social context may be as discriminable before hearing a phee stimulus as during or immediately after the stimulus. Also, coefficients in PC1 span positive and negative values, indicating that some frontal neurons have greater firing rates for the antiphonal context, whereas others are more active for the independent context. Importantly, PC coefficients do not distinguish contributions to variance between social contexts (i.e., context separation) from contributions to variance within social contexts, and so these implications must be verified with direct tests.

Initially, we sought to test these implications by measuring the accuracy of social-context classification using a two-means classifier that takes advantage of the separation between antiphonal and independent stimuli in PC1 (see Materials and Methods, Two-means classification). This classifier performed well for test simulations of frontal cortical activity from

large neuron populations, but not for individual units. Figure 1.3A shows two distributions that illustrate classification accuracy for our entire population of units (magenta histogram) compared with the units individually (gray histogram). Median accuracy for individual units was 51%, only slightly better than chance performance of 50% correct, though this was highly significant (signed-rank test, $z = 4.97$, $p < 0.0001$, $n = 258$ units), and even the best individual unit classified stimulus context with only 72% accuracy (Monte Carlo cross-validation, $p < 0.002$). In contrast, median accuracy for the entire population of units was 91%, significantly greater than the most accurate single unit (Monte Carlo cross-validation, $p < 0.002$). This indicates that the variance in PC1 used to classify neural activity emerges from the large population of units, once again indicating that many units likely help distinguish between social contexts. It is also possible that population classification may benefit from the methods required to simulate responses.

When simulating the frontal cortex population responses, activity across neurons is decorrelated because all units were not recorded in the same behavioral session. This is shown in Figure 1.3B, which estimates the distribution of pairwise correlations of all frontal units from the only the frontal units simultaneously recorded within a behavioral session. When phee-stimulus responses were maintained across all units (normal), median pairwise correlations were 0.11. When unit responses to stimuli were shuffled within each of the social contexts (shuffled, as occurs for the population simulations), median pairwise correlations decreased to 0.05 (signed-rank test, $z = 11.9$, $p < 0.0001$, $n = 258$ units). To address how this might affect the population classifier within the constraints of our data, we compared accuracies for each session before (normal) and after shuffling (shuffled) responses within social contexts (Fig. 3C). Sessions typically had few units (median of four), which resulted in most accuracies only slightly above chance, similar to individual units (Fig. 3A). Nevertheless, median accuracy increased by ~0.01

when responses were shuffled, a proportional increase by ~40% above chance (signed-rank test, $z = -2.15$, $p = 0.032$, $n = 62$ sessions). Removing this proportional improvement from our population classifier (91% median accuracy, 41% above chance) results in a median accuracy of 77%, which still performs significantly better than the median accuracy of the best individual unit (Monte Carlo cross-validation, largest $p = 0.014$).

We also tested whether activity from each of the four stimulus time periods (Pre, Voc 1, Voc 2, and Post) could identify phee-stimulus social context using the same population classifier using unit activity only in the respective time period. The accuracy of classification is given, along with 95% CIs calculated from 500 simulations, in Figure 1.3D, in which all four time periods show significant accuracies well above chance. Despite its limitations, our classifier illustrates the power of small activity changes in large neuronal populations in determining context. Next, we applied simpler analyses to measure social context-dependent changes in individual units and across time periods.

We examined two sample units with high PC1 coefficient magnitudes as exemplars to guide further analysis. Figure 1.4A shows an example raster plot of unit activity from one behavioral session (top) summarized by normalized firing rates in 0.5 s time bins (bottom). This example unit corresponded to large positive PC1 coefficients, which, based on the initial population analysis, is expected to be more active for antiphonal phee stimuli. While this trend is apparent before, during, and after stimuli are heard, which is consistent with the PC1 coefficients in each time period, the raster plot shows substantial variability, and a low enough firing rate that differences within 0.5 s time bins are rarely significant. Figure 1.4B shows the activity of an example unit with large negative PC1 coefficients, displayed in the same format as Figure 1.4A. In this example, firing rates tend to favor independent phee stimuli. Also, as in Figure 1.4A, this

example exhibits this preference before, during, and after stimuli are heard, but again, comparisons rarely reach significance over the 0.5 s time bins. From an examination of these particular units, it seems the difficulty in finding significant changes in activity across contexts has to do with the low firing rates of these frontal units engaging in these natural vocal exchanges. Figure 1.4 also plots the mean activity for each single unit (Fig. 4C; 172 of 258) and each multiunit (Fig. 4D; 86 of 258), averaged over all time periods for the antiphonal context compared with the independent context. Typically, changes in activity were <1 Hz; however, these changes could be quite large as a proportion of their mean firing rates (mean of 2.5 Hz for single units and 3.3 Hz for multiunits), with a mean difference between contexts of 10% for single units and 18% for multiunits. Averaging over longer time periods, or across many units, could reveal significant differences despite the low firing rates.

We quantified the prevalence of social-context response preferences, as observed in the example units above, by calculating a context preference index for each unit spanning all four stimulus time periods (Pre, Voc 1, Voc 2, and Post; see Materials and Methods, Determining stimulus preference for individual units). Of the all 258 units, 43 (17%) significantly distinguished between social contexts (Monte Carlo cross-validation, $p < 0.05$), and 155 (60%) had a positive context preference index (signed-rank test, $z = 6.48$, $p < 0.0001$, $n = 258$ units). Figure 1.5A shows the context preference index of each unit, with blue indicating antiphonal preferring units and red for independent. Notably, preference is almost evenly split, both for units with significant preferences (40% antiphonal to 60% independent) and over all units (43% antiphonal to 57% independent). Eliminating the firing-rate normalizations revealed an average unit change in firing rate between preferred and nonpreferred contexts is quite small (mean, <1 Hz), making analyses of individual units at finer time scales impractical. The context preference

index may miss important units that show interactions between social context and the phee-stimulus periods. For example, unit 247 from Figure 1.5A may play such a role. It has large negative PC1 coefficients during Voc 1 and Voc 2 but a large positive coefficient during the Post time period (Fig. 2C), and yet the context preference index is negative. Notably, the context preference index is strongly correlated with the unit PC1 coefficient magnitudes ($r(256) = 0.80$, $p < 0.0001$), illustrated in Figure 1.5A by ordering units by increasing coefficient magnitudes, validating the use of the coefficients for identifying sources of variance between social contexts.

Because most units (60%) had a consistent phee preference, we tested whether the entire population of units could distinguish between stimulus contexts on a finer time scale. Figure 1.5B plots the mean normalized firing rates of all 258 units for preferred stimuli compared with nonpreferred stimuli; as in Figure 1.5A, the data used to determine the preferred context was omitted. Firing rates were significantly different at every time point from 1.5 s before phee onset to 6 s after ($t(257) < -3.3$, $p < 0.001$, all points remain significant after Holm–Sidak correction for multiple comparisons). Notably, this shows differences in activity between social contexts of phee stimuli before they are even heard. To confirm that our analyses for Figure 1.5A,B were unbiased, they were performed after randomly shuffling the social context assigned to each stimulus and for stimuli categorized by phee length (Fig. 5C–F). Neither controls reached significance, with fewer individual units showing significant differences than expected by chance (4.7 and 3%, binomial test, $p = 1$ and 0.20, respectively) and no significant differences in population activity in any time period [$t(257) < 1.77$ (magnitude), $p > 0.089$]. In summary, we find that a substantial proportion of individual units in the frontal cortex differentiate between the social context of vocalizations when responses are averaged over several seconds, and the

combined activity of many frontal units distinguish the social context on finer time scales, even before the stimulus is heard.

In addition to changes in firing rate, we also tested for differences in interneuronal correlations associated with stimulus social context (see Materials and Methods, Measurement of neuronal correlations). We estimated the average magnitude of pairwise correlations for units recorded in the same behavioral sessions separately for each social context, but otherwise using the same methods as in Figure 1.3B. The population of units had median interneuronal correlations of 0.12 for antiphonal stimuli compared with 0.09 for independent stimuli (signed-rank test, $z = 6.5$, $p < 0.0001$, $n = 258$ units). Thus, in addition to changes in frontal cortex firing rates, interneuronal correlations are also greater within the antiphonal social context.

In the analyses performed above, we included cortical units from all four arrays to increase the power of our analyses. It is possible that several of our results are only possible when combining all units or that only distinct areas of the frontal cortex exhibit different changes in unit activity. However, we recorded nonoverlapping populations of neurons throughout marmoset areas 6, 8, 45, and 47 in the frontal cortex from four electrode arrays in three different subjects. The positions of each array are illustrated in Figure 1.6. We found no obvious indication that anatomical location corresponded to the context preference index of units, except that Array B02 exhibited the weakest preferences. Array B02 also included the fewest units ($n = 28$ units; <11%). We averaged activity across units from each array using the same methods as in Figure 1.5B except that we used longer time windows (specifically the Pre, Voc 1, Voc 2, and Post time periods) and we also combined all time periods. We found that Arrays B01, F01, and R01 all had significant differences in activity across one or more time periods, and all were significant for the Pre period ($t(118,68,46) = -2.17, -2.68, -2.02$, $p = 0.034, 0.009, 0.046$,

respectively, no correction for multiple comparisons) and for all time periods combined ($t(118,68,46) = -2.29, -2.30, -2.68$, $p = 0.027, 0.026, 0.010$, respectively, no correction for multiple comparisons). Only Array B02 did not show consistent significant differences. This suggests that the role of the frontal cortex in distinguishing between phee contexts is not limited to one area, although the extent throughout all of the frontal cortex remains unknown. Also, by analyzing each array separately, we confirm that our results are reproducible in all three subjects.

Antiphonal conversations in marmosets are characterized by the reciprocal exchange of vocalizations (Fig. 1). In the final set of analyses, we investigated how neural activity was affected by sequences of phee stimuli within these conversations, rather than the individual instances of independent and antiphonal stimuli targeted in all previous analyses. We refer to consecutive sequences of stimuli within a single context as "bouts," with conversations occurring during antiphonal bouts. To compare activity during bouts, we calculated a population response, which averages activity across units by normalizing and rectifying stimulus spike rates (see Materials and Methods, Bout categorization). Figure 1.7A shows that unit activity is tightly coupled to social context. Repeated-measures ANOVA found significant interaction between bout category and the position in the bout ($F(5,1480) = 7.915$, $p = 0.005$). Population responses significantly change between the end of an independent bout and the start of an antiphonal conversation ($p < 0.005$, Tukey's range test, $df = 2313$, $\alpha = 0.05$). This difference in activity persists over the course of the conversation and reliably changes again. The response to the first independent stimulus does not reach significance compared with the final stimulus of an antiphonal bout, but the responses to subsequent independent stimuli are significantly different ($p < 0.0354$, Tukey's range test, $df = 2313$, $\alpha = 0.05$). This pattern emphasizes that the behavioral outcome is closely coupled with a change in firing rate across the population. Notably, there is

no difference in the stimuli at the time they are broadcast, yet the latter exhibits the shift in neural activity even before the stimulus presentation. In other words, although the first antiphonal stimulus in a conversation is not deemed antiphonal until the subject produces a response several seconds later, the change in firing rate is evident before the vocalization is heard and persists over the length of the conversation. This has occurred because, presumably, the state of the frontal cortex has shifted to mediate conversations.

There is some indication that the bout length may affect neural firing rates, though data are limited. Using the same normalization method as used in the previous bout analysis, Figure 1.7B plots population responses for all antiphonal and independent stimuli across the population, as well as those that occurred in bouts of 1, 2, 3, 4, 5, or more phee stimuli. In general, there is a trend toward more extreme responses over longer sequences of independent stimuli with significant difference reached at the ≥5 bout length ($p < 0.03$, t test, Holm–Bonferroni corrected $df = 55$, $\alpha = 0.05$). A similar trend is evident for antiphonal stimuli, but too few long conversations were available to convincingly determine this case. A two-way ANOVA test of stimulus context and bout length shows significant interaction and group mean differences ($p < 0.001$, F test, $df = 4$, $\alpha = 0.05$). These analyses suggest a linear relationship between neural activity and the length of the natural conversation.

1.5 Discussion

We examined the activity of frontal cortical neurons recorded from areas 6, 8, 45, and 47 of freely moving marmoset subjects engaged in natural vocal conversations with a virtual marmoset to characterize how neural activity distinguished between two social contexts in which phee calls are heard. Namely, occasions when a phee elicits a conspecific vocal response (antiphonal context) and those that do not (independent context). We found small (∼1 Hz), but

widespread, changes in activity across neural populations within all frontal areas sampled. Notably, this population of units did not tend to exhibit stimulus-driven responses to hearing vocalizations produced by conspecifics. In fact, the period before stimulus onset was comparable to periods during or after the phee stimulus in the degree to which the population activity distinguished between the two social contexts (Figs. 3D, 5B). Finally, not only was a robust correlation evident between frontal cortex activity and antiphonal conversations (Fig. 7A), but the magnitude of the neural response increased as a function of conversation length (Fig. 7B), supporting the notion that this neuronal process is strongly related to the social context of these natural vocal exchanges. It is possible that the magnitude of the change in the neural activity at the time the conversation initiated determined its eventual duration, or it could be that these changes became increasingly affected as the conversation persisted, potentially due to neuronal coupling that may occur between both individuals during the vocal interaction (Stephens et al., 2010; Hasson et al., 2012; Silbert et al., 2014). These important facets of frontal activity help narrow the potential role of this activity in the process of natural communication.

The pattern of activity observed here is particularly notable given the constraints imposed on neurophysiological recordings of the frontal cortex in freely moving, naturally behaving monkeys. Although the overall effect was most clearly evident when pooling activity across the population, 17% of individual units showed significant differences between the antiphonal and independent social contexts (Fig. 5A). This number likely underestimates the proportion of units with changes in activity related to social context because many units showed different patterns of activity across the time periods before, during, and after phee stimuli based on PCA (Fig. 2C). This type of response complexity likely contributes to the accurate classification of social context (91%; Fig. 3A), which substantially outperformed results from a reasonably comparable

study in which classification of conditioned auditory task behavior was based on prefrontal neuron activity (Russ et al., 2008). Furthermore, of units in which we observed a significant difference, slightly fewer units showed activity preferences in the antiphonal context (40%) compared with the independent context (60%), suggesting that the temporal epoch of each unit is not only where heterogeneity of the population occurs, but is also where preference for a particular social context is evident. One notable difficulty with regards to our analysis was the small changes observed in firing rate. We are, however, highly confident that these changes are significant, because no differences were evident when the same analyses were applied to randomly assigned social contexts or the classification of phee-stimulus length (Fig. 2). These analyses paint an intricate picture in which multiple mechanisms may support the observed pattern of response, potentially in coordination with a broader process critical to primate social communication that will only manifest under natural conditions.

Many processes are involved in active social signaling, including sensory processing, recognition, categorization, decision making, attention, and arousal. A key question for the current study is which mechanism, or more likely mechanisms, may underlie the observed changes in frontal cortex activity during natural marmoset conversations, and which may not. First, general wakefulness can be ruled out as a key contributing factor because animals were monitored continuously, and remained awake throughout these recordings. General arousal from stress is also unlikely. The marmosets were habituated to the experimental setup and exhibited no overt signs of stress. Also, sensory-driven processes, such as encoding the phee stimulus or decision making based upon the phee are unlikely because differences in neural activity were comparable in magnitude before, during, and after the phee stimuli were broadcast.

The frontal cortex activity reported here is likely related to some facet of attention and/or arousal, which are both often poorly defined terms that can refer to a wide range of mechanisms (Harris and Thiele, 2011). Each are also likely synonymous with nearly all active primate social behaviors, and difficult to disambiguate in natural contexts. With regards to selective attention of sensory information, attentive states show reduced neuronal noise correlations (Cohen and Maunsell, 2009; Mitchell et al., 2009; Harris and Thiele, 2011), which is notably different from the increase in unit correlations that we observe in the engaged, antiphonal, context. Moreover, it does not resemble the known mechanism for selective attention, which corresponds to large changes in neuronal activity localized to specific frontal cortical areas (Gregoriou et al., 2012). It seems more likely that if this activity is related to attention, it would be more related to a concept of "sustained attention" (Sarter and Bruno, 2000), which is not well distinguished from aspects of arousal. Given behavioral evidence showing that marmosets acutely attend to the behavior of multiple individuals during antiphonal conversations (Toarmino et al., 2017b) and coordinate the timing of these exchanges based on the behavior of conspecifics (Roy et al., 2011), it is reasonable to assume that some type of attentional mechanisms contribute to the pattern of activity reported here in the frontal cortex.

It is also probable that a broad variety of processes referred to as arousal may have modulated frontal cortex responses during natural conversations (McGinley et al., 2015). With regards to the sensory cortex, arousal refers to multiple behavioral states, some of which have similar effects on sensory processing. Key among them is desynchronization of neural activity, which can help sensory encoding, and increased activity in particular types of neurons (McGinley et al., 2015; Vinck et al., 2015). Remarkably few studies, however, have observed the mechanisms of such arousal in the frontal cortex, and none in a naturally behaving primate. In a

socially engaged antiphonal state, we observed fewer units with increased activity than those with decreased activity (Fig. 5A), and also a greater degree of interneuronal correlations. In this case, a broad, correlated, and distinct pattern of frontal activity could shift marmosets between levels of social arousal or receptiveness. Likewise, small changes in firing may also be ideal for maintaining the behavioral state with neuromodulators, such as acetylcholine, which is associated with various types of arousal (McKenna et al., 1989; Sarter and Bruno, 2000). As a result, individual firing rates across the population, even on the order of 1 Hz reported, could have substantial influence on behavior, especially when they are more tightly correlated and persist over several seconds, such as during antiphonal conversations.

Marmoset antiphonal conversations are characterized by the reciprocal, coordinated exchange of vocalizations between conspecifics. We hypothesize that the observed change in frontal cortex activity indicates a shift in brain state that facilitates social monitoring, a process critical to natural human and nonhuman primate social interactions, including conversations. While this type of shift in behavioral state cannot account for the full complexity of natural conversations, one key characteristic of this human and nonhuman primate behavior is coordinated turn-taking, in which individuals alternate speaking and listening (Levinson, 2016). To produce an appropriate response, an individual must attend to a conspecific ongoing behavior while suppressing their own motor behavior. The change in the state of the frontal cortex may reflect a change in social arousal and attention, and serve a sensory gating function to facilitate rapid processing of conspecific vocalizations throughout the auditory system (Miller et al., 2010b; Petkov et al., 2015) and precipitate the cascade of subsequent social decision-making processes (Toarmino et al., 2017a). The observed neuronal process could also enable neuronal coupling to improve the communicative efficacy of the conversations, similar to what has been

shown in human fMRI experiments (Stephens et al., 2010; Silbert et al., 2014). Because these experiments involved a marmoset engaging in conversations with a virtual marmoset, rather than a live marmoset, we cannot test this latter hypothesis, which will be a key target in future studies.

Primate sociality is somewhat paradoxical. Whereas primate social cognition is dynamic and sophisticated, the content and number of social signals is relatively limited despite their fundamental role in mediating these complex social interactions. Resolving this contradiction may necessitate understanding not only what individual social signals communicate but also how they are used within the myriad of ongoing social interactions that typify primate societies. The approach taken here offers unique opportunities to investigate communication within the dynamic, natural contexts that more fully encapsulate the myriad of neural mechanisms that support primate sociality. Neuronal processes, such as the social context-dependent change in frontal cortex state reported here, may occur only when primates are actively interacting with each other. Considerations of how these active dimensions of communication unfold over time within the context of natural primate social life may lead to unique insights into the intricate complexities of the primate social brain.

1.6 Acknowledgements

This work supported by the National Institutes of Health Grant R01 DC012087 to C.T.M.

Chapter 1, in full, is a reprint of the material as it appears in Social Context-Dependent Activity in Marmoset Frontal Cortex Populations during Natural Conversations 2017. Nummela, Samuel U.; Jovanovic, Vladimir; Miller, Cory T.; de la Mothe, Lisa, Journal of Neuroscience, 2017. The dissertation author was one of the primary investigators and authors of this paper.

1.7 Figures

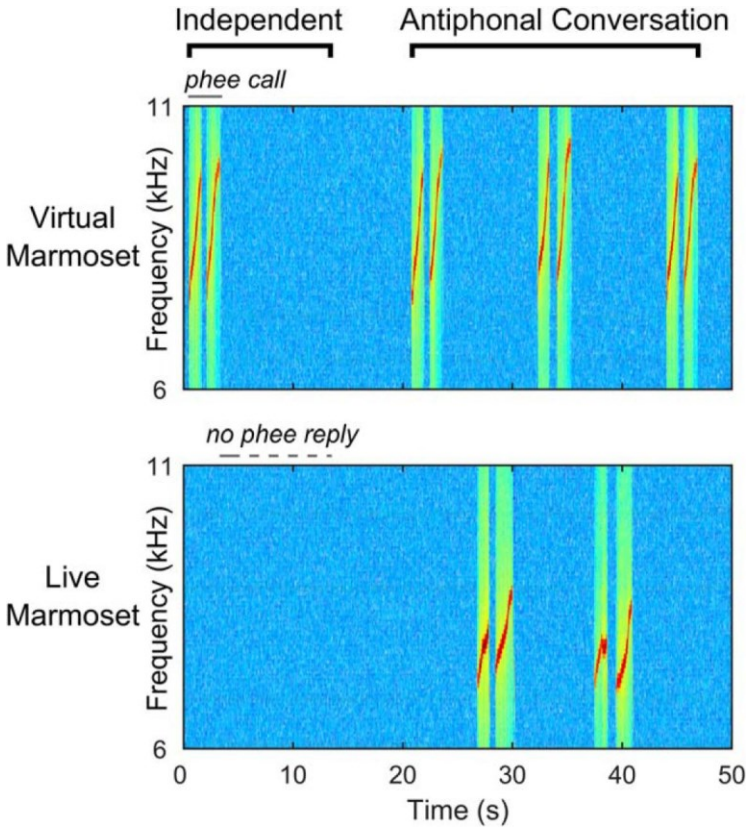


Figure 1.1: Antiphonal conversations in marmosets. Spectrograms of antiphonal and independent phee calls. Top, the virtual marmoset phee stimuli broadcast to the marmoset subject. Bottom, Phee calls from the live marmoset Subject M. The first virtual marmoset phee call is an independent stimulus, characterized by the absence of response from M within the antiphonal period of 10 s as denoted by a gray dashed line. The next two calls from the virtual marmoset are antiphonal stimuli, characterized by phee responses from M within the antiphonal period. The final call from the virtual marmoset is independent, with no vocal response from M within 10 s.

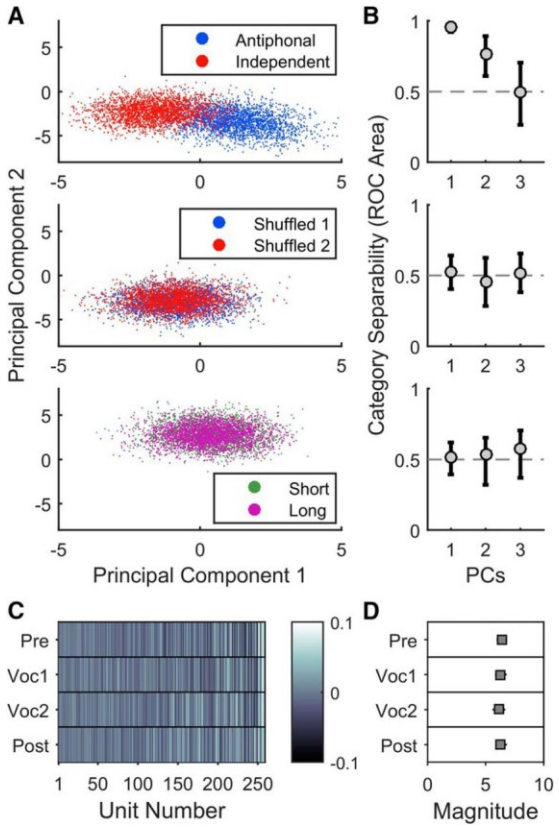


Figure 1.2: Frontal cortical activity separates vocalization social contexts. (A) Sample frontal cortical population responses simulated from test datasets plotted in the first and second PCs from training simulations by phee social context (top), randomly shuffled contexts (middle), and phee-stimulus lengths (bottom). The population responses to phee social contexts form distinct clusters for antiphonal and independent contexts, but this is not the case for randomly shuffled contexts, or to stimuli separated by phee length. (B) ROC analysis measures the separation of population responses by social context (top), randomly shuffled contexts (middle), and stimulus length (bottom), in the first three PCs from independent training datasets. An area under the ROC curve of 0.5 indicates stimulus categories are not separable, and 1 indicates they are completely separable. Population responses to independent and antiphonal phee calls are highly separable in PC1, and remain significantly separable in the PC2. The remaining PCs show no separation. Population responses to randomly shuffled contexts (middle) or to different phee stimulus lengths (bottom) are not separable. Error bars are 95% CIs. (C) The median coefficients of PC1 from all 500 training simulations of population responses to antiphonal and independent stimuli. Coefficients are organized by unit (columns) and time period (rows). Units are sorted by the sum of the coefficient magnitudes over all time periods, such that recording 1 contributes the least to PC1 and recording 258 contributes the most. (D) Mean and 95% CIs for median PC1 coefficient contributions from each time period, calculated by summing the coefficient magnitudes across all units. Error bars are 95% CIs.

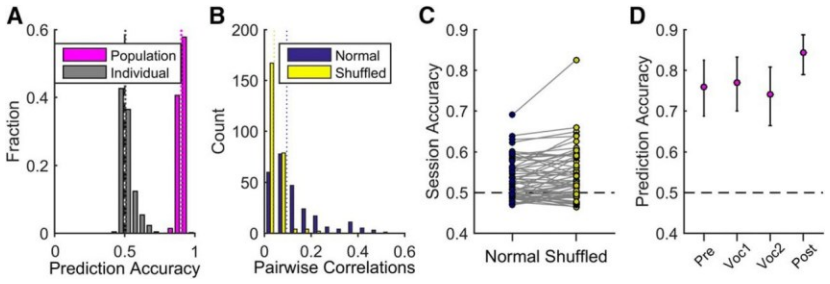


Figure 1.3: Social context classification from PC1 emerges from the population activity. (A) Histograms of individual unit classifier accuracies and the distribution of accuracies of the population classifier performed on all 500 population response simulations. (B) The distribution of average pairwise correlation coefficients over all units estimated under two conditions: with responses to each phee stimulus maintained across all units within a session (normal) and with responses to phee stimuli shuffled within each context across units (shuffled). (C) The change in classifier accuracy for each behavioral session with responses to each stimulus maintained across all units in that session (normal) and with responses to each stimulus shuffled, within social context, across units (shuffled). (D) Accuracy of the population classifier is much greater than chance even when predicting stimulus context from population activity at only one time period relative to that stimulus. Error bars are 95% CIs.

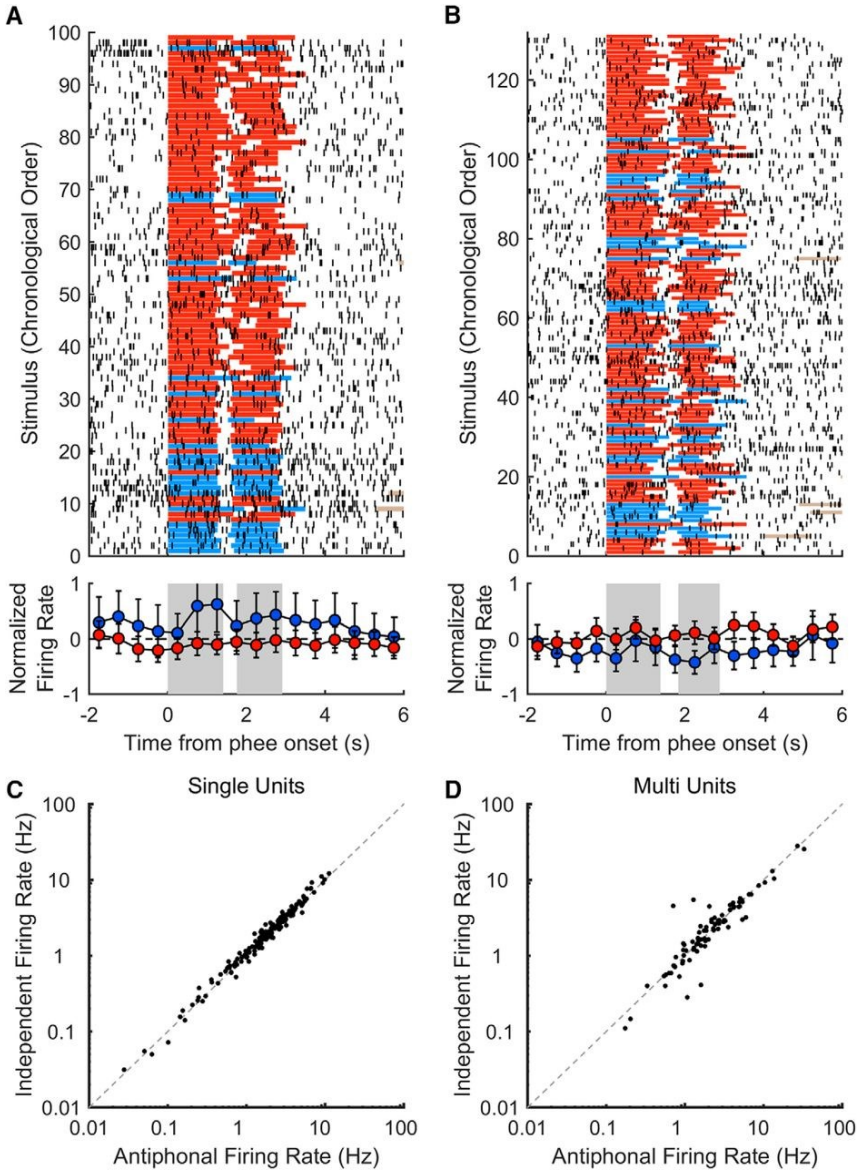


Figure 1.4: Differences in unit activity between vocalization social contexts. (A) Sample raster (1 ms resolution; top) and normalized spike rates (0.5 s time bins; bottom) for a unit with large positive PC1 coefficients. In the raster plot, red lines indicate independent phee stimuli, blue lines indicate antiphonal phee stimuli, and brown lines mark subject replies (when within the axis limits). Binned, normalized, firing rates are shown below, with blue points for antiphonal stimuli and red points for independent stimuli. Gray rectangles indicate the mean phee pulse times. Error bars are 95% CIs. (B) Sample raster with same conventions as A, except for a recording with preference for independent stimuli, which had large negative PC1 coefficients. (C) Mean firing rates of all 172 single units in response to antiphonal stimuli compared with independent stimuli. Firing rates were averaged over all four time periods and plotted on a logarithmic scale. (D) Mean firing rates of all 86 multiunits using the same conventions as C.

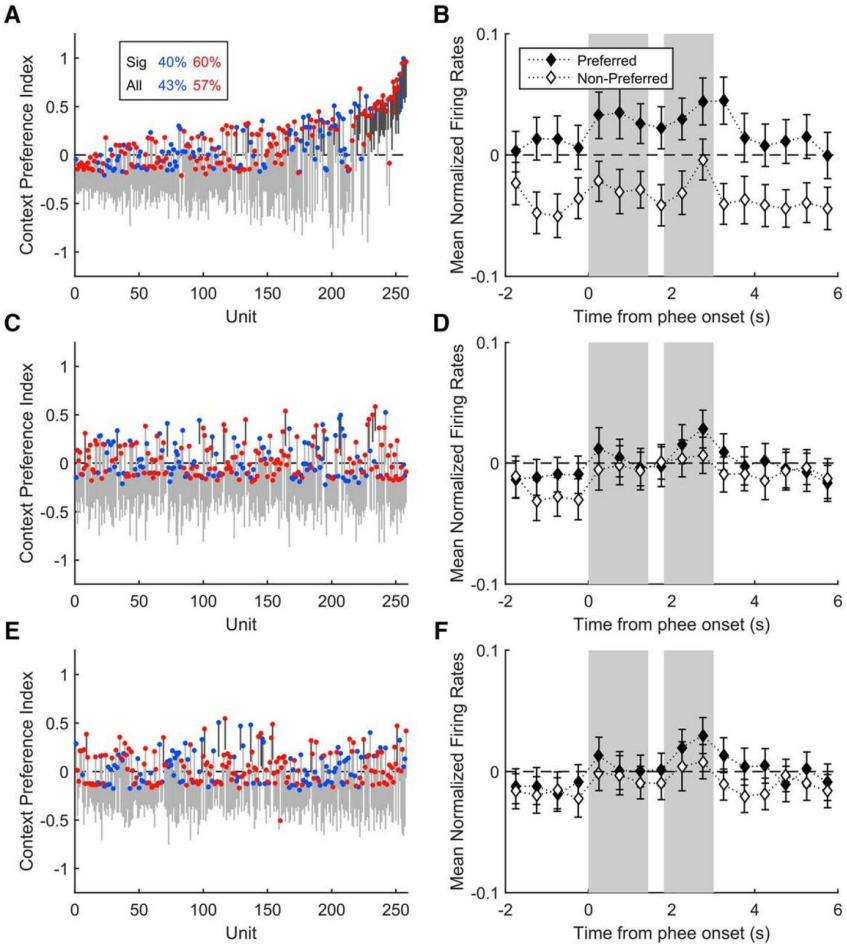


Figure 1.5: Frontal cortical activity distinguishes between vocalization social contexts. (A) The context preference index, given by the difference in mean normalized firing rate between preferred and nonpreferred stimuli contexts, is plotted for each unit. Blue points indicate preference for the antiphonal context and red points for the independent context. Error bars are one-tailed 95% CIs. The inset provides the percentage of units preferring each social context, indicated by color, for the units that reached significance (Sig) and for all units (All). The example units from Figure 1.4A,B have respectively colored error bars. (B) Mean normalized firing rates over all units for the preferred phee context (black) and nonpreferred phee context (white), in 0.5 s time bins. Error bars are 95% CIs. Mean phee-stimulus pulse timings are indicated by gray rectangles. (C, D) Same conventions as A and B except context preference index was calculated when phee contexts were randomly assigned by shuffling context identity. (E, F) Same conventions as A and B except context preference index was calculated based on length of the phee stimulus instead of its social context. In C and E, unit positions were kept the same as in A, with colors corresponding to the phee context preferences of each unit in A.

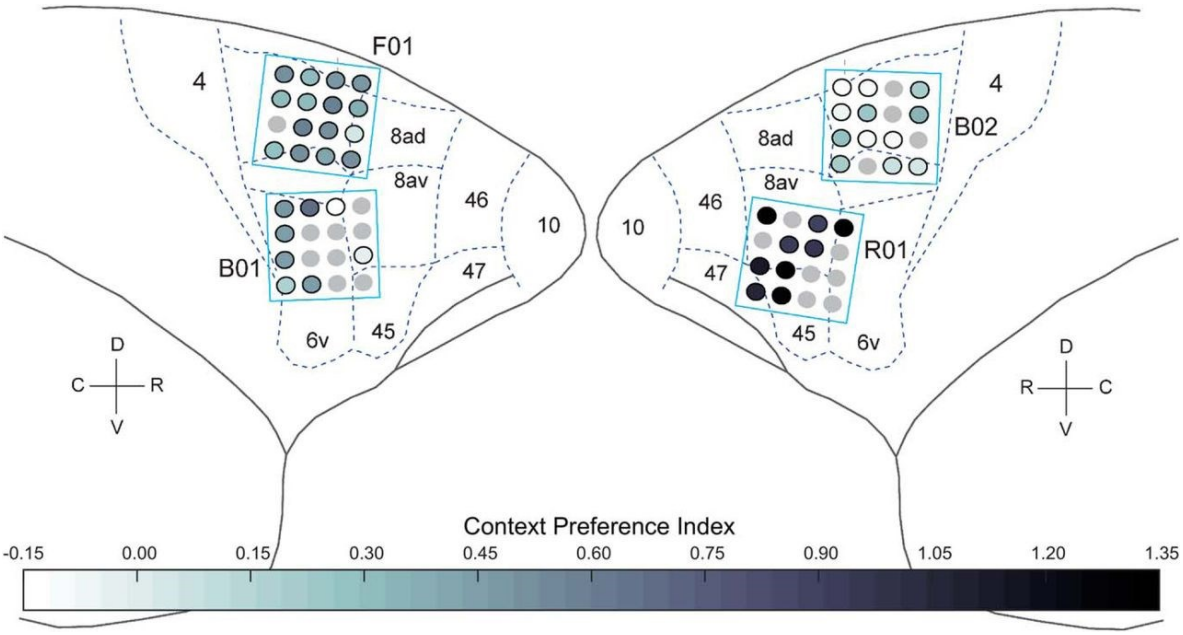


Figure 1.6: Discrimination of social contexts by location of electrode arrays in the frontal cortex. The anatomical layout of the four electrode arrays are shown. B01 and B02 are arrays from Subject B, and F01 and R01 are from Subjects F and R. Each electrode is colored according to its context preference index. Channels with multiple units only show the highest value; channels with no units are light gray with no border.

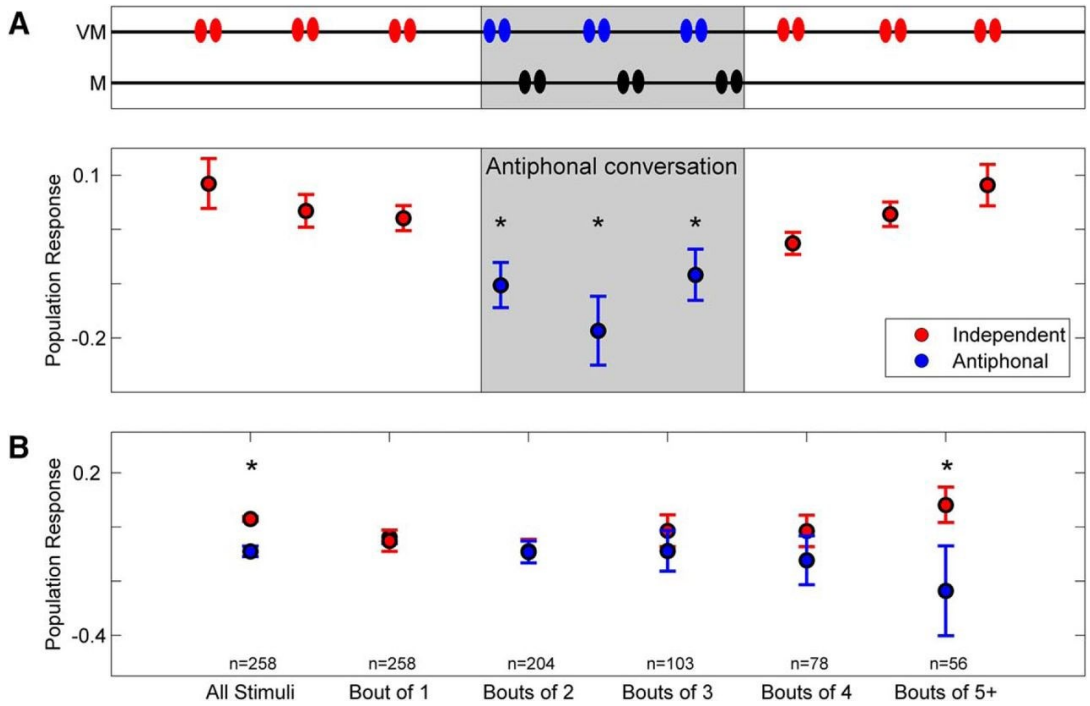


Figure 1.7: Population responses during conversations. (A) A schematic of an antiphonal conversation between a virtual monkey (VM) and marmoset subject (M) showing a bout of three independent stimuli (red), followed by a conversational bout of three antiphonal stimuli (blue) with subject replies (black), and then another independent bout. Below, Neural population responses to stimuli within the conversation. Responses to phee stimuli at the end and start of independent bouts (red) are greater than responses to stimuli at the start, middle, and end of an antiphonal conversation (blue). *$p < 0.05$ differences for antiphonal bouts compared with independent bouts. (B) Population responses to antiphonal (blue) and independent (red) stimuli are compared with responses during antiphonal and independent bouts of specified lengths. Below each comparison n is the number of units with data for the bout length. *$p < 0.03$ for differences between independent and antiphonal contexts. All error bars are 95% Cis.

1.8 References

Cheney DL, Seyfarth RM (2007) Baboon metaphysics: the evolution of a social mind. Chicago: University of Chicago.Google Scholar

Chow C, Mitchell J, Miller CT (2015) Vocal turn-taking in a nonhuman primate is learned during ontogeny. Proc R Soc B 282:210150069. doi:10.1098/rspb.2015.0069

Cohen MR, Maunsell JH (2009) Attention improves performance primarily by reducing interneuronal correlation. Nat Neurosci 12:1594–1600. doi:10.1038/nn.2439 pmid:19915566

Dunbar RIM (2003) The social brain: mind, language and society in evolutionary perspective. Annu Rev Anthropol 32:163–181. doi:10.1146/annurev.anthro.32.061002.093158

Eliades SJ, Miller CT (2017) Marmoset vocal communication: neurobiology and behavior. Dev Neurobiol 77:286–299. doi:10.1002/dneu.22464 pmid:27739195

Eliades SJ, Wang X (2008a) Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature 453:1102–1106. doi:10.1038/nature06910 pmid:18454135

Eliades SJ, Wang X (2008b) Chronic multi-electrode neural recording in free-roaming monkeys. J Neurosci Methods 172:201–214. doi:10.1016/j.jneumeth.2008.04.029 pmid:18572250

Engh AL, Hoffeier RR, Cheney DL, Seyfarth RM (2006) Who, me? Can baboons infer the target of vocalizations? Anim Behav 71:381–387. doi:10.1016/j.anbehav.2005.05.009

Fisher C, Freiwald WA (2015) Whole-agent selectivity within the macaque face processing system. Proc Natl Acad Sci U S A 112:14717–14722. doi:10.1073/pnas.1512378112 pmid:26464511

Gifford GW 3rd., MacLean KA, Hauser MD, Cohen YE (2005) The neurophysiology of functionally meaningful categories: macaque ventrolateral prefrontal cortex plays a critical role in spontaneous categorization of species-specific vocalizations. J Cogn Neurosci 17:1471–1482. doi:10.1162/0898929054985464 pmid:16197700

Gregoriou GG, Gotts SJ, Desimone R (2012) Cell-type-specific synchronization of neural activity in FEF with V4 during attention. Neuron 73:581–594. doi:10.1016/j.neuron.2011.12.019 pmid:22325208

Harris KD, Thiele A (2011) Cortical state and attention. Nat Rev Neurosci 12:509–523. doi:10.1038/nrn3084 pmid:21829219

Hasson U, Ghazanfar AA, Galantucci B, Garrod S, Keysers C (2012) Brain-to-brain coupling: a mechanism for creating and sharing a social world. Trends Cogn Sci 16:114–121. doi:10.1016/j.tics.2011.12.007 pmid:22221820

Hickok G, Poeppel D (2004) Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92:67–99. doi:10.1016/j.cognition.2003.10.011 pmid:15037127

Hung CC, Yen CC, Ciuchta JL, Papoti D, Bock NA, Leopold DA, Silva AC (2015) Functional mapping of face-selective regions in the extrastriate visual cortex of the monkey. J Neurosci 35:1160–1172. doi:10.1523/JNEUROSCI.2659-14.2015 pmid:25609630

Leopold DA, Bondar IV, Giese MA (2006) Norm-based face encoding by single neurons in the monkey inferotemporal cortex. Nature 442:572–575. doi:10.1038/nature04951 pmid:16862123

Levinson SC (2016) Turn-taking in human communication: origins and implications for language processing. Trends Cogn Sci 20:6–14. doi:10.1016/j.tics.2015.10.010 pmid:26651245

McGinley MJ, Vinck M, Reimer J, Batista-Brito R, Zagha E, Cadwell CR, Tolias AS, Cardin JA, McCormick DA (2015) Waking state: rapid variations modulate neural and behavioral responses. Neuron 87:1143–1161. doi:10.1016/j.neuron.2015.09.012 pmid:26402600

McKenna TM, Ashe JH, Weinberger NM (1989) Cholinergic modulation of frequency receptive fields in auditory cortex: I. Frequency-specific effects of muscarinic agonists. Synapse 4:30–43. doi:10.1002/syn.890040105 pmid:2672402

Miller CT, Wang X (2006) Sensory-motor interactions modulate a primate vocal behavior: antiphonal calling in common marmosets. J Comp Physiol A Neuroethol Sens Neural Behav Physiol 192:27–38. doi:10.1007/s00359-005-0043-z pmid:16133500

Miller CT, Wren Thomas A (2012) Individual recognition during bouts of antiphonal calling in common marmosets. J Comp Physiol A Neuroethol Sens Neural Behav Physiol 198:337–346. doi:10.1007/s00359-012-0712-7 pmid:22277952

Miller CT, Beck K, Meade B, Wang X (2009) Antiphonal call timing in marmosets is behaviorally significant: interactive playback experiments. J Comp Physiol A Neuroethol Sens Neural Behav Physiol 195:783–789. doi:10.1007/s00359-009-0456-1 pmid:19597736

Miller CT, Mandel K, Wang X (2010a) The communicative content of the common marmoset phee call during antiphonal calling. Am J Primatol 72:974–980. doi:10.1002/ajp.20854 pmid:20549761

Miller CT, Dimauro A, Pistorio A, Hendry S, Wang X (2010b) Vocalization induced cFos expression in marmoset cortex. Front Integr Neurosci 4:128. doi:10.3389/fnint.2010.00128 pmid:21179582

Miller CT, Thomas AW, Nummela SU, de la Mothe LA (2015) Responses of primate frontal cortex neurons during natural vocal communication. J Neurophysiol 114:1158–1171. doi:10.1152/jn.01003.2014 pmid:26084912

Miller CT, Freiwald WA, Leopold DA, Mitchell JF, Silva AC, Wang X (2016) Marmosets: a neuroscientific model of human social behavior. Neuron 90:219–233. doi:10.1016/j.neuron.2016.03.018 pmid:27100195

Mitchell JF, Sundberg KA, Reynolds JH (2009) Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. Neuron 63:879–888. doi:10.1016/j.neuron.2009.09.013 pmid:19778515

Newman JD, Lindsley DF (1976) Single unit analysis of auditory processing in squirrel monkey frontal cortex. Exp Brain Res 25:169–181. pmid:819284

Perrodin C, Kayser C, Logothetis NK, Petkov CI (2011) Voice cells in primate temporal lobe. Curr Biol 21:1408–1415. doi:10.1016/j.cub.2011.07.028 pmid:21835625

Petkov CI, Kikuchi Y, Milne AE, Mishkin M, Rauschecker JP, Logothetis NK (2015) Different forms of effective connectivity in primate frontotemporal pathways. Nat Commun 6:6000. doi:10.1038/ncomms7000 pmid:25613079

Platt ML, Seyfarth RM, Cheney DL (2016) Adaptations for social cognition in the primate brain. Philos Trans R Soc Lond B Biol Sci 371:20150096. doi:10.1098/rstb.2015.0096 pmid:26729935

Romanski LM, Averbeck BB, Diltz M (2005) Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. J Neurophysiol 93:734–747. pmid:15371495

Rosati A, Santos L, Hare B (2010) Primate social cognition: thirty years after Premack and Woodruff. In: Primate neuroethology (Platt M, Ghazanfar AA, eds), pp 117–142. Oxford, UK: Oxford UP.

Roy S, Miller CT, Gottsch D, Wang X (2011) Vocal control by the common marmoset in the presence of interfering noise. J Exp Biol 214:3619–3629. doi:10.1242/jeb.056101 pmid:21993791

Roy S, Zhao L, Wang X (2016) Distinct neural activities in premotor cortex during natural vocal behaviors in a New World primate, the common marmoset (Callithrix jacchus). J Neurosci 36:12168–12179. doi:10.1523/JNEUROSCI.1646-16.2016 pmid:27903726

Russ BE, Orr LE, Cohen YE (2008) Prefrontal neurons predict choices during an auditory same-different task. Curr Biol 18:1483–1488. doi:10.1016/j.cub.2008.08.054 pmid:18818080

Sarter M, Bruno JP (2000) Cortical cholinertic inputs mediating arousal, attentional processing and dreaming: differential afferent regulation of the basal forebrain by telencephalic and brainstem afferents. Neuroscience 95:933–952. pmid:10682701

Schiel N, Souto A (2017) The common marmoset: an overview of its natural history, ecology and behavior. Dev Neurobiol 77:244–262. doi:10.1002/dneu.22458 pmid:27706919

Seyfarth RM, Cheney DL (2014) Evolution of language from social cognition. Curr Op Neurobiol 28:5–9. doi:10.1016/j.conb.2014.04.003 pmid:24813180

Silbert LJ, Honey CJ, Simony E, Poeppel D, Hasson U (2014) Coupled neural systems underlie the production comprehension of naturalistic narrative speech. Proc Natl Acad Sci U S A 111:E4687–E4696. pmid:25267658

Stephens GJ, Silbert LJ, Hasson U (2010) Speaker-listener neural coupling underlies successful communication. Proc Natl Acad Sci U S A 107:14425–14430. doi:10.1073/pnas.1008662107 pmid:20660768

Toarmino CR, Jovanovic V, Miller CT (2017a) Decisions to communicate in the primate ecological and social landscapes. In: Psychological mechanisms in animal communication (Bee MA, Miller CT, eds), pp 271–284. New York: Springer.

Toarmino CR, Wong L, Miller CT (2017b) Audience affects decision-making in a marmoset communication network. Biol Lett 13:pii:20160934. doi:10.1098/rsbl.2016.0934 pmid:28100720

Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. Science 311:670–674. doi:10.1126/science.1119983 pmid:16456083

Vinck M, Batista-Brito R, Knoblich U, Cardin JA (2015) Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. Neuron 86:740–754. doi:10.1016/j.neuron.2015.03.028 pmid:25892300.

2 Within-neuron comparison illustrates context-dependence of natural social signal processing in primate prefrontal cortex

2.1 Abstract

Communication is an inherently interactive process involving the exchange of social signals between conspecifics that is heavily affected by the nuances of the social contexts. Yet the neural basis of primate communication has primarily been studied in head-restrained paradigms in which social signals are entirely divorced from the contexts in which they naturally occur. The few studies to examine the neural basis of vocal communication in freely-moving monkeys have yielded patterns of results that diverge notably from parallel experiments using more traditional approaches. Here we sought to directly test – within neuron – whether these observations reflect differences in experimental design or more fundamental insight into how the primate brain functions under natural conditions. We recorded from the responses of the same neurons to acoustic stimuli – including vocalizations - across a series of contexts ranging from passive-listening in head-restrained to freely-moving monkeys engaged in interactive conversational exchanges. After examining within-neuron differences, we found that passive listening across any condition could elicit the robust responses to vocalizations and that subjects' mobility did not significantly affect neural response, but neural responses when subjects were actively engaged in communication were typically weaker or absent entirely. However, PFC activity during natural communication was modulated by meaningful events, such as conversational exchanges. We investigated this observation explicitly by selectively changing the expected vocal stimulus during these interactions. In those single meaningful events, a population of PFC neurons exhibited robust neural responses further suggesting that the dynamics of this key neural structure are strongly influenced by the nuances of natural

communication behaviors. More broadly, these results suggest that neural responses to social signals in more restrictive contexts are not predictive of how those same neurons respond under more naturalistic contexts, at least in PFC. Furthermore, these findings highlight the importance of naturalistic experimental environments when trying to study complex behaviors in animal models.

2.2 Introduction.

Brain evolution occurs through selection acting directly on behaviors, inescapably coupling the supporting neural circuits and their phenotypic manifestations as two sides of the same coin (Briscoe and Ragsdale, 2018; Miller et al., 2019). This intricate relationship affords a powerful vehicle to interrogate neural circuits by leveraging the naturally occurring behaviors they evolved to support. Indeed, this approach has been widely used to elucidate characteristics of brain function in a range of different species, including several lines of work that have resulted in Nobel Prizes (Yartsev, 2017). By contrast, studies of primate neural circuits have been largely divorced from the species-typical behaviors they evolved to support, instead biasing to experimental designs involving restrained animals passively presented with stimuli and conditioned behavioral paradigms. While undoubtedly highly productive, this conceptual framework rests on the key assumption that results derived from these highly controlled, reductionistic experiments reveal facets of neural function that are also employed under more naturalistic contexts. An assumption that it seems may not be true, or at least not ubiquitously true (McMahon et al., 2015). Here we sought to test this issue by directly comparing the responses of individual neurons in marmoset prefrontal cortex to vocalizations and other acoustic stimuli across a series of contexts ranging from the head-restrained paradigm commonly used in primate neuroscience research to freely-moving monkeys engaged in natural communication.

We posited if experimental contexts had limited effects on the nature of neuronal activity, the predictive value of data collected in more traditional contexts for understanding natural brain function would be high. If, however, the converse occurred, and neural responses during the traditional context were consistently dissimilar to natural behavior, it would suggest that the predictive value would be far more limited.

The disparity between natural behaviors and the traditional approaches used to examine these processes is perhaps most evident in studies of social communication. By its nature, communication is an inherently interactive process involving the exchange of signals between conspecifics (Guilford and Dawkins, 1991; Hauser, 1996). Under natural conditions, these signals function to mediate social interactions between conspecifics within the backdrop of dynamic primate social landscapes that is rich with nuanced contextual information (Toarmino et al., 2017b). Yet social signals – such as faces and vocalizations – are routinely presented as static stimuli entirely devoid of any social context. Certainly, modern approaches have been prolific, revealing integrated networks for both face and voice processing within the primate brain (Tsao et al., 2006; Petkov et al., 2008; Tsao and Livingstone, 2008; Perrodin et al., 2011; Freiwald et al., 2016). The rationale behind this framework is consistent with the broader conceptions that simple, reductionistic approaches reveal the foundational principles of primate brain function. However, the only study to address this assumption did not yield supporting evidence. McMahon and colleagues (McMahon et al., 2015) found that neurons in the AF face patch exhibited classic selectivity to face stimuli when subjects passively viewed static images of faces, but the responses of the same neurons were driven by entirely different features when subjects viewed videos of monkeys engaged in social interactions. That this relatively small contextual change resulted in such dramatic changes within individually identified face cells belies the deeper

question of what changes would occur if subjects were directly engaged in a social interaction, rather than as a passive observer to interactions.

The lateral prefrontal cortex is a neocortical substrate unique to primates that has been attributed as a key supporting a myriad of higher cognitive processes (Miller and Cohen, 2001; Fuster, 2008), including vocal communication. Neurophysiological studies of PFC neurons in head-restrained macaque monkeys passively presented with vocalizations reported that populations of neurons were responsive to conspecific calls (Gifford et al., 2005; Romanski et al., 2005; Averbeck and Romanski, 2006; Romanski and Averbeck, 2009; Plakke et al., 2013a). By contrast, experiments in which prefrontal and premotor neurons were recorded while marmoset monkeys engaged in their natural vocal interactions found that these cells did not exhibit a stimulus driven response when hearing conspecific vocalizations (Miller et al., 2015). The notable disparity between these findings offers an ideal scenario in which to test the effects of behavioral context on primate brain function and the relationship between neuronal activity in both traditional and naturalistic contexts. Indeed, a recent study revealed that neurons in marmoset frontal cortex exhibited small, but reliable, changes in firing rate when subjects heard a conspecific vocalization that almost perfectly predicted their propensity to engage in a vocal interaction, suggesting that the state of the neural population upon hearing determined the monkey's propensity to socially engage (Nummela et al., 2017). Taken together, this series of experiments suggests that social context may significantly affect patterns of neural activity in primate frontal cortex. Here we sought to test this hypothesis by comparing the responses of single neurons in marmoset prefrontal cortex to vocalizations across a series of contexts – ranging from a more traditional, head-restrained paradigm to freely-moving monkeys engaged in natural communication. By directly comparing responses within an individual neuron across

these contexts, we aimed to ascertain how factors, including subject's mobility and stimulus presentation, affected patterns of neural activity in prefrontal cortex.

2.3 Methods.

2.3.1 Subjects

Five adult common marmosets (Callithrix jacchus) were used for this experiment. A01 and H02 were male while M01, E01, and H01 were female. All subjects were at least 1.5 years old at time of implant. M01 had two arrays implanted on the left hemisphere: one rostral to area 8av and another at the temporal gyrus. Only the PFC array was used for all analysis mentioned here. E01 had bilateral arrays implanted frontal cortex in a similar area to M01. A01 had a bilateral implant of arrays as well, but only one array was viable (right rostral to 8av). H01 had a left hemisphere PFC implant, while H02 had a right hemisphere. In sum, 5 arrays had usable data. Arrays from E01 and A01 had the worst difficulty with grounding and represented the smallest set of units used. Precise locations of the arrays will not be covered in this chapter. The histology and subsequent analyses have been delayed due to a centennial pandemic. The locations of the arrays were designed to be rostral to previous work showing weak neural responses to vocal signals during natural communication in areas 8av and 45 (Miller et al., 2015). All animals were group housed, and experiments were performed in the Cortical Systems and Behavior Laboratory at University of California San Diego (UCSD). Experiments were approved by the UCSD Institutional Animal Care and Use Committee.

2.3.2 Behavioral Paradigm

The recording sessions included three main experimental conditions, or contexts, for the subjects. The three conditions were presented in randomized order each session to prevent any ordering effect between contexts. Briefly outlined from a more detailed description in prior work

(Miller et al. 2015), there was an Interactive context wherein the subject engaged with a Virtual Monkey (VM) who played back socially relevant "Antiphonal" (phees produced in response to another phee) and "Independent" calls (phees produced in absence of conspecific call and spontaneously). This session was kept the same as in prior work except for an introduction of Probe stimuli. Every two to three conversational exchanges of antiphonal calls between subject and VM, the VM had a 50% chance to play a phee call from another monkey (same or different sex) or a reversal of its phee call (Figure 2.9A).

Two other contexts were presented to the monkey that were both passive-listening tasks. In one, Restrained, the subject would be chair- and head- restrained while facing forward towards the occluder. In Freely context, the subject was free to move around within a clear acrylic box as outlined in recording procedures. In both cases, subjects would hear blocks of phee stimuli and broadband noise (M01, E01, A01). H01 and H02 also had twitter calls, reverse twitter, and reverse phee calls introduced. The stimuli would be played in four blocks of vocalizations and noise, twice each. The order was randomized on every recording session to prevent any ordering effect. Vocalizations with H01 and H02 were shuffled across all examples. In all cases, each category of vocalization and noise were played 30 times each. M01, E01, and A01 had an ISI of 1 second and 3 second for each block type. H01 and H02 only had 1 second ISI across all blocks due to no differences found in unit activity between those the two ISI durations. Between each block, 30 seconds of silence occurred, and both contexts started and ended with five minutes of silence to get a baseline of the unit activity.

H01 and H02 had a final condition to determine whether Noise driven neurons exhibited spatial tuning (referred to as Spatial context). Subjects would be head-/chair- restrained and presented 30 broadband noise stimuli with 30 seconds in between each block and 30 seconds

before and after. Subjects' chairs would be randomly reoriented on a center point of the table. With 0° facing the speaker, the monkey would start on a random 45° position offset from forward and systematically move through all eight directions in a random sequence for each recording session.

2.3.3 Behavioral Recording Procedures

All recording sessions occurred in a Radio-Frequency shielding room (ETS-Lindgren) in a 4 x 3 m room. On opposing narrow sides of the room were the speaker (far) and subject (near). Both were placed on tables with a black cloth occluder in the middle of the room to prevent the subject from seeing the speaker. Subjects were placed in a clear acrylic box with a plastic mesh on the front side facing the occluder side (32 x 18 x 46 cm). One microphone was placed in front of the subject and speaker each (Sennheiser, model ME-66). The speaker broadcast the phees and noise at an approximate 80-90 dbSPL measured 1 m from the speaker (Polk Audio TSi100, frequency range 40-22,000 Hz).

Subject and speaker calls were recorded simultaneously along with the neurophysiological data on the same data acquisition card (NI PCI-6254). Subject data was analyzed midstream via MATLAB during Interactive contexts to determine whether the subject made a call. Calls produced by the monkey that hit an acoustic range from 6 to 10 kHz and were at least 1 second long were classified as phee calls. A phee call by the subject within 10 seconds of a VM phee call was classified as a response or antiphonal phee call by the subject. Calls outside that range were classified as spontaneous. The computer always responded to subject calls within 2 to 3 seconds, or made an Independent call every 45 to 90 sec.

For H01 and H02, we measured subject head position during the Freely and Interactive contexts. Two cameras (GoPro Hero Session) faced the right side and top side of the clear box

relative to the front side facing the occluder and speaker. During these recordings, an Arduino system flashed an LED visible to both cameras at 0.5 Hz. This signal was recorded along with the audio and neural streams as well which allowed us to accurately align the video streams with the start of neural recording. Images at stimulus onset were grabbed from the top-down viewing camera, and subject head position was noted relative to the front. This allowed us to see if there was any orientation selectivity of freely moving animals.

2.3.4 Neurophysiological Recording Procedures

After 1.5 years of age, subjects were implanted with acrylic head caps with stainless steel head posts. During the surgery, the lateral sulcus was marked, and the rostral and lateral edges of the frontal cortex were visible. We used these markings to determine the locations of the microelectrode arrays. The arrays were not MRI compatible, and thus histological analysis will be performed on each subjects' brain to provide precise placement of all electrodes. Due to a pandemic, there have been significant delays in that regard.

The microelectrode arrays were 16 channel Warp16 electrode arrays (Neuralynx). Each array had 16 independent guide tubes in a 4 x 4 mm grid with tungsten electrodes. Each array is implanted on the surface of the brain making each electrode within the tubes enter the laminar tissue perpendicularly when pushed by a Warp Drive pusher. The calibrated Warp Drive pusher would attach to the end of a guide tube to allow advancement of 10 to 20 μm per electrode twice a week.

Electrodes were recorded at 20,000 Hz with a prefilter at 1 Hz to 9000 Hz, and 20,000 gain. Subjects had 1:1 gain, headstage preamplifier attached to the Warp16 arrays that was attached to a sufficiently long tether to allow subjects to freely move around in the box. A metal coil tightly wrapped around the tether prevented any interference by the subject during Freely

and Interactive contexts. Offline spike sorting was done by hand by combining across multiple sessions recorded in a single day, applying a 300 to 9000 Hz filter and thresholding subsamples across the entire recording session. Units with at least 10 dB SNR were included, with a <1% interspike intervals <1 msec refractory period. Overall, 400 isolated single units were found, with some channels collecting multiple well isolated single units (Figure 2.1A). As can be seen, the thresholding applied easily discovered the three single units and PCA clustering shows the discrepancy in statistical structure of each waveform (Figure 2.1B) and also showing how they were held across the duration of the recording sessions (typically lasting about a 100 minutes for H01 and H02, and 90 min for the rest of the subjects). Overall, the units maintained their structure across the multiple session. Of the 400 single units, 200 units were maintained across all three conditions. Typical drop-off of the remaining 200 units occurred when electrode channels would have an introduced noise far exceeding the threshold set in another session. This usually occurred going from Restrained to a Freely/Interactive contexts and occasionally from Freely/Interactive to Restrained contexts. Future work will include auto-sorting the units and including multi-unit level analysis.

2.3.5 Data Analysis

2.3.5.1 Unit Significance.

A unit had significant response to a stimulus if it satisfied one of these conditions at α=0.05 significance level with a Sign Rank test: (1) the 1000 msec prior to onset of trials was significantly less than the 1000 msec after onset, (2) the prior was significantly less than the firing rate during the entire duration of calls (only applied to phees), and, finally, (3) 500 msec prior was significantly less than 500 msec around the timing of the peak in the units PSTH during each trial.

2.3.5.2 Spatial-Tuned Units.

For the Orientation context outlined in the Behavioral Paradigm subsection, all 240 Noise trials were used in a 2-Way ANOVA. The 500 msec before and after onset were used for comparison across all trials. And each trial was subgrouped according to the direction the subject was facing. If the unit had a significant main effect on before and after, it was included for further analysis. The interactive effect was then tested and Tukey-Kramer corrected comparisons were used to determine if there was any orientation that had significance compared to the others. If the unit had no orientation selectivity or all orientations were selective with a main effect for before and after, we counted that unit as being generally responsive (all 8 directions). In practice, only 1 unit had significant selectivity for each direction as well as overall.

In the Freely and Interactive contexts, a similar analysis was used for Noise and Phee. As noted in Behavioral Recording Procedures, the orientation of the front of the head relative to the speaker on the transverse plane was marked. These orientations were binned into eight groups for comparison to the Orientation context. In practice, most of the time, the subjects were looking at 3 general directions. Even so, we included any that had at least 5 same stimuli for a given bin. 2-Way ANOVA did not find any units responsive to a given orientation.

2.3.5.3 PSTH Normalization.

Unit trials were binned at 100 msec intervals starting 1000 msec prior to onset and extending 1500 msec after onset for Noise, Twitter, and Reverse Twitter. Phee and Reverse Phees extended 4000 msec. The firing rate for each bin within each trial for a given unit was calculated. Normalization occurred by taking all the bins prior to onset to get the mean and standard deviation. The Z-score was calculated based off of those values. For any given unit, we could then find the mean across each bin to get a single normalized PSTH line. When multiple

units are shown together, they typically have the mean PSTH from each combined together with standard error bars around them. In some cases, like in the Significant Single Event, all trials are combined to get the average response as each individual unit is not contributing its entire set of trials.

2.3.5.4 Firing Rate Normalization.

To compare Firing Rate changes from Restrained to Freely contexts, we took the 1000 msec prior to onset and 1000 msec after onset. The prior Firing Rates were used to Z-score the subsequent mean after Firing Rates.

2.3.5.5 Classification.

Each unit included in the classification had to have at least 5 events from each class of interest. If it did not, it would not be included in any part of the simulations. 1000 Monte-Carlo simulations were created by subsampling each unit's class of trials 2250 times with replacement for training and test sets. The trials would be split in half between test and training in random permutation each simulation. Each trial would have 9 points of data (1 bin for 500 msec before, 1 for 500 msec after a Phee call, 3 bins for each of the 2 pulses, and 1 bin for the inter-pulse-interval). In sum, there would be 2250 randomly chosen training trials for each class, and the equivalent in test trials. The 2250 x 9 matrices for each unit would then be combined with all other units with sufficient trials, for each of the classes within both training and test. For example, 32 units in IFR for Restrained, Freely, and Interactive classes would produce a matrix of size 6750 x 288 for both training and test with no overlap of trials across those two. Each row is then assigned a value representing the class that all 288 values represent.

The training data was fit using the "fitcecoc" function within MATLAB's using the default settings. This creates multiclass support vector machines for each simulation, and then

predicts the class of the test data. Each confusion matrix that results from the predictions is then stored, and the average value of those matrices is used in Figure 2.6. We also quantified the performance of each classifier by the Matthews Correlation Coefficient (MCC). For a K x K confusion matrix C (i.e. K = 3 for Restrained, Freely, and Interactive classes in Figure 2.6C):

$$MCC = \frac{\sum_k \sum_l \sum_m C_{kk} C_{lm} - C_{kl} C_{mk}}{\sqrt{\sum_k (\sum_l C_{kl})(\sum_{k'|k' \neq k} \sum_{l'} C_{k'l'})} \sqrt{\sum_k (\sum_l C_{lk})(\sum_{k'|k' \neq k} \sum_{l'} C_{l'k'})}}$$

For K = 2 (Figure 2.6E), MCC ranges from -1 to +1. The +1 means perfect prediction, -1 means complete disagreement between predicted and actual, and 0 is no better than chance prediction. For K = 3 (Figure 2.6B), the lower limit is not at -1 and unique to any given classifier. Still, 0 means the chance predictions and +1 is perfect prediction. We conducted a null-hypothesis test for both classifiers running the same size data as the full data set (all 200 units). With a randomized assignment of class for each row in training and test data, the MCC was at 0 as expected.

2.3.5.6 Significant Single Events.

To determine if a single event was significant, we took the trials and from 300 msec prior to onset and 4300 msec post onset (only Phee calls were analyzed). We then binned by 300 msec intervals with 150 msec overlap. The lone bin prior to onset for all relevant trials was used to z-score all bins for that unit's trial data. If any trial had a z-score of 2 or more (i.e. 2 standard deviations above the mean of the onset), the trial was tagged as a significant single event. The ratio of each trial above the 2 SD was also noted. Thus, we took all the trials across all the units maintained across the three conditions. For the heatmaps seen in Figure 2.7 and 8, we combined all significant trials across all units and interpolated 100 units for each trial to give a smooth transition and calculate the intermediate time points as well.

2.3.5.7 Conversation Lengths.

For Figure 2.8, we looked at the various conversation lengths and the effect of significant single events within them. We categorized three different conversation lengths: No response (0-1 calls by subject), Short conversation (2-3 calls by subject), Long conversation (4+ calls by subject). If the subject only makes one call in response to a VM call, it was included in No Response. Any Independent calls that were not responded to were also included. Notably, the Antiphonal call after a monkey response was not included if there was not a subsequent response by the subject. Short conversations would require the monkey respond for each subsequent VM call. The first call could be Independent, but subsequent VM calls would be Antiphonal. Long conversations lumped any conversation with 4 or more exchanges (and thus 4 or more VM calls responded to by the subject). For the purpose of determining continuous conversations, we included all VM calls including Probe and Control calls as the subject may or may not respond to those.

2.4 Results.

2.4.1 Stimuli Responsiveness

We first analyzed the overall neural responsiveness to the different acoustic presented to subjects. Figure 2.2A shows the response characteristics of single units that exhibited a significant change in neural activity in response to each of the acoustic stimuli - Twitter, Phee, their Reversals, and Noise - in the Restrained and/or Freely conditions. The normalized PSTH was calculated for each unit that exhibiting a significant response to a given stimulus and context combination. The normalized PSTH of each responsive unit was collapsed into a mean normalized PSTH with 95% confidence intervals. As can be seen, the units had modulated response to the Twitter and Phees as well as Reverse Twitter and Reverse Phees and Noise

stimuli. Overall, analyses showed that 400 single units (~ 85%) exhibited a statistically significant change in activity in response to at least one acoustic stimulus in at least one test context. When looking at those single units only within Restrained and Freely, we find that the responsiveness rate varies across stimulus sets but not necessarily contexts. Figure 2.2B shows the percentage of responding neurons to each stimulus set in a given context. For Twitter calls, Phee calls, their reversals, and Noise, there was comparable number of neurons responding in both the Restrained and Freely context suggesting that under identical stimulus presentation paradigms, mobility did not reduce the likelihood of PFC responses to acoustic stimuli. Figure 2.2C outlines the area of interest that we implanted arrays. Future anatomical work will address the precise locations.

2.4.2 Impact of Mobility

In the next set of analyses, we examined the responses of individual neurons that were recorded i both the Restrained and Freely contexts. Notably, we broadcast the identical stimuli in the identical stimulus presentation pattern – vocalizations and noise – to subjects in each of these contexts. The only difference being subjects mobility, as they were freely-moving when hearing the stimuli in the Freely context. We found 256 units for the Noise condition and 205 units for the Phee condition. The phee condition had a lower number of units due to misplaying phees in M01. Figure 2.3A & B shows the change in mean firing rate prior to onset of a stimulus and during the duration of the stimulus. For each trial in a given set, the firing rate during the duration of the stimulus was normalized to the 500 msec prior to the onset of the stimulus. We observed more Noise responsive neurons exhibited a stronger firing rate during the restrained than in Freely contexts (81/121 units, 66.94%), while the pattern was more evenly distributed for Phee responsive neurons (27/52 units, 51.92%). Figure 2.3C outlines the overall percentage of

each category of responsiveness (Restrained, Freely, or Both). For Noise, the vast majority of neurons responded to both contexts (121/257) and much less so for Restrained (24/257) or Free (27/257). For Phee, there was a fair distribution across all three categories with Both (52/206) still having the most compared to the Restrained (41/206) or Freely (43/206). For Figure 2.3D, we took the same set of units that had a response to Noise or Phee in both contexts and found that that the general trend of increased variance in the Freely condition relative to Restrained (Noise: 97/121 units, 80.17%, Phee: 36/52, 69.23%). Thus, we find that a majority of the units had a weaker response (mean Firing Rate) in the Freely condition, but also had a higher variance (mean Standard Deviation of Firing Rate).

As mentioned in the Methods section, we also recorded the subject orientation in the Freely and Interactive contexts to determine whether subjects' head orientation relative to the speaker impacted each units response properties. For those two subjects, we also had a stimulus presentation of noise bursts while their restrained position was systematically rotated during Noise presentations. These two conditions allowed us to test whether there was any effect on head direction (binned into 45° arcs) relative to the source of the stimulus that might be driving some of the responses we saw in Figure 2.3A & B. For the Freely condition, 0 of 103 units had a significant change in neural activity in response to the stimulus to any of the eight head directions classified for either Noise or Phee calls. Furthermore, Figure 2.3E plots the number of spatial bins for which a unit exhibited a significant response. We found 74 units that were responsive to the Noise stimulus played in the rotating context (74/156 units, 47.44%). Of those exhibiting a response, only 2 showed any spatial selectivity (2.70%) while the remaining 72 neurons were responsive to noise stimuli equally in all in all eight conditions (97.3%). Taken together, the head direction of the subject relative to the speaker does not appear to be a

significant source of variance for PFC responsiveness to acoustic stimuli, further casting doubt on that mobility plays a substantive role in how this population represents these sounds.

2.4.3 Impact of Interactivity

We next sought to explicate the impact of stimulus presentation pattern on PFC neuron responses. While mobility did affect moderate changes in neural responses to the acoustic stimuli, the Interactive context sought to more directly contrast passive presentation of acoustic stimuli with vocal signal processing during natural communication. We analyzed the 200 well isolated neurons that were held stable across all three contexts: Restrained, Freely, and Interactive. As described in the Methods, the Interactive condition involved subjects engaging in natural vocal interaction with a Virtual Monkey in order to instigate the naturally occurring reciprocal conversational exchanges of antiphonal phee calls (Miller et al., 2009; Miller and Thomas, 2012; Toarmino et al., 2017a).

Overall, we observed that 151 PFC neurons (75.5%) exhibited a significant change in activity to phee calls in at least one of the three contexts. These neurons were placed in one of 8 possible categories based on the contexts in which they exhibited significant stimulus responses: Restrained only, Freely only, Interactive only, Restrained and Freely only, Restrained and Interactive only, Freely and Interactive only, and Restrained, Freely and Interactive. Figure 2.4A plots the mean normalized PSTH for all units found within the 8 possible categories. The grey lines plot the mean firing rate of contexts for which the units exhibited no significant stimulus response, while the colored lines represent the context(s) for which stimulus response was statistically significant. As can be seen, aside from the F R and I F R groups, significant responses were characterized by a consistent, but modest change in neural activity. By contrast, neural responses the F R and I F R category neurons exhibited robust changes in neural activity

that were entrained to the temporal structure of the phee call itself. Figure 2.4B plots the mean normalized PSTH for each context in which units had a significant response towards. Once again, the weakest response is seen in the Interactive condition across units that have any responsiveness

Figure 2.5A and Figure 2.5B show exemplar single units representative of the two categories of neurons that had strongly driven responses. Figure 2.5A shows a unit that has a significant response to phees in the Restrained, Freely, and not the Interactive context despite subjects hearing the identical phee calls in all three contexts. The neuron shown in Figure 2.5B exhibited a different pattern. This neuron exhibited a response in all three contexts, and we can safely assume that the unit was not affected by the context that the phee stimulus was presented to the subject.

Figure 2.6A shows the mean normalized response for Restrained and Freely responding units and demonstrates that the response to each pulse in the phee stimuli had similar responses. Figure 2.6B shows units that had preference for all three contexts, and in those there is a clear bump in the second pulse. Of note is the Interactive context that has a much stronger response to the first pulse than the second pulse but in general elicited a more modest response compared to the other two contexts. We further unpacked these data by distinguishing between two contexts that occurred during the Interactive experiments. Specifically, in the Interactive context the VM broadcast phees both in response to the subject's own phee call (Antiphonal calls) and spontaneously after long periods of silence by the subject to evoke a vocal response from subjects (Independent calls) (Nummela et al., 2017). The mean normalized PSTH for I F R with the Interactive context separated by whether the phee was presented in the Antiphonal or Independent context is shown in Figure 2.6C. By separating the Antiphonal and Independent

events, we observed that the response seen in Restrained and Freely was comparable to the Independent context. By contrast, phee calls presented in the antiphonal context elicited a modest change in neural activity. This suggests that the class of neurons that exhibited strong phee responses to phees across the three main experiment contexts, IFR neurons, were also significantly affected by nuances of the social interactions and whether subjects were actively engaged in a communication exchange (Antiphonal) or passively listening to phee calls (Independent).

2.4.4 Classifying Contexts

The next set of analyses directly compared the pattern of neural responses in each of the three contexts to determine similarities and differences. As a first step, we examined the distribution of the neural responses in each context by displaying the normalized FR for each trial across all units in normal probability plots that compares the actual distribution for the given context to the normal distribution (Figure 2.7A). None of the three contexts adhered to a normal distribution and had a right skew. Given the three contexts, we tested the distributions against each other to determine statistical similarity using the Two-sample Kolmogorov-Smirnov test with a Bonferroni correction on the significance value (α=0.05/6, 0.0083). While the Restrained and Freely contexts were statistically similar, showing no difference in their pattern, the Interactive distribution was significantly different from both the Restrained and Freely contexts.

We next examined how differentiable the three distributions were by testing how well we could classify between the three contexts. As outlined in the Methods section, each trial was divided into 9 bins (1 before, after, in between pulses, and 3 for each pulse). Each value was normalized to the firing rate prior to the onset of the trials. We ran 1000 simulations across three different data sets: only I F R responding units (IFR: 32 units), any responding units (Any: 151

units), or all units found (All: 200 units). ANOVA found that there was a difference in MCC, a classification performance metric, across the three contexts (One-way ANOVA, df = 2998, p = 0.000). Furthermore, as shown in Figure 2.7B, the IFR had the worst performance and was significantly different from the other two (Tukey-Kramer corrected, IFR*Any & IFR*All p =0.000, Any*All p=0.962). Figure 2.7C plots the mean confusion matrix for all 1000 simulations. For IFR, the classification system had difficulty differentiating between the Restrained and Freely contexts but not Interactive with large percentages misclassified to each other's cases. The Interactive context meanwhile is likewise misclassified as Freely but not to the same degree as the two other classes were misclassified to it. Finally, in the best performing data sets, there was little false positives of Interactive cases (1.0% towards Freely). Yet, we still see that at least 6% of Freely or Restrained cases are misclassified to the other. This suggests that there are real differences in the overall structure of response at a trial level basis for the Interactive conditions with respect to Freely and Restrained.

Figure 2.7D plots the distribution of normalized Firing Rate within the Interactive context, once more distinguishing between the Antiphonal and Independent call cases. Given these two contexts, we once more applied the Two-sample Kolmogorov-Smirnov test with a Bonferroni correction on the significance value (α=0.05/2, 0.025), but the distributions were not significantly different (p = 0.344). We next ran the 1000 simulations as described above, using IFR units, Any, and All units (Figure 2.7E). Once more, we found that there was a difference in performance across the three data sets (One-way ANOVA, df = 2998, p = 0.000). When quantifying the multiple comparisons, we found that there was significant difference across all three data sets (Tukey-Kramer corrected, IFR*Any & IFR*All & Any*All p = 0.000). Figure 2.7F plots the mean performance by confusion matrix across the three data sets. What is notable

is the trend to have a higher rate of False Positives marking true Antiphonal as predicted Independent (8.4%) compared to true Independent to predicted Antiphonal (3.7%). Overall, analyses show that there are differences in the responses to Antiphonal calls versus Independent calls. Likewise, we see that those calls as a set (Interactive) have a response significantly different from Restrained or Freely contexts within the same units. The fact that there was some similarity to the Antiphonal classes to Independent suggests that there may be some trials within Antiphonal that are similar to Independent.

2.4.5 Single Event Analysis

From the classification tests, we found that including all acoustic responsive single units that were stable throughout the three contexts performed significantly better than the IFR responsive units alone. We also observed that just taking All Unit data performed better in classifying across Antiphonal and Independent calls. This suggests that there is some degree of coding for these phee calls across many of the neurons, even if individually their change in firing rate did not exceed statistical thresholds. To explore this further, we next investigated neural activity for single events – single phee stimulus presentation - that individually exceeded a firing rate threshold as outlined in the Methods. Figure 2.8A shows a heatmap for all significant events (defined as having at least one bin above 2 SD to the bin prior to onset for trials within the unit). Each trial was interpolated across 100 points to the bin prior to the onset of the call to 4.3 seconds after onset of the trial. This allowed us to make comparisons and bin all the values as appropriate before and after the phee call, the first and second pulse, and the gap between the two pulses. As can be seen across the three conditions, there is a similar modulated average response to the bins for Restrained and Freely but less so for Interactive. Furthermore, Figure 2.8B shows that there were less significant single events in a unit in Interactive versus the two

other two conditions (1-Way ANOVA, df = 607, p = 0.001. Tukey-Kramer corrected I < R p = 0.001, I < F p = 0.03). As well in Figure 2.8C, the ratio of bins in a given single event greater than 2 SD was significantly less in the Interactive condition (1-Way ANOVA, df = 9720, p = 0.000. Tukey-Kramer corrected I < R p = 0.000, I < F p = 0.000). This suggests that Interactive contexts have less significant events than expected and spends less time than expected on average above a significant threshold, which explains why there is typically only modest changes in Interactive context.

PFC responses during the Antiphonal and Independent trials exhibited more tepid responses that were driven by the stimulus onset but not as robustly as seen in Restrained or Freely (Figure 2.8D). Unlike the overall PSTH responses of Independent units in Figure 2.6C, Independent single events were dissimilar to both the Restrained and Freely contexts suggesting that the significant single events that occur within that context are dissimilar at the single event level. By using only normalized PSTH plots for comparison, this would have been missed and speaks to the usefulness of single event analysis. Figure 2.8E shows the ratio of significant single events to other trials within each unit, revealing no statistical difference (1-Way ANOVA, df = 305, p = 0.861). Figure 2.8F plots the ratio of events above the 2 SD threshold for those significant single events revealing a significantly higher ration in the Independent context relative to the Antiphonal context (1-Way ANOVA, df = 3532, p = 0.004). Overall, we can see that phee calls heard in the Antiphonal context exhibited the weakest response out of all sets of trials (Restrained, Freely, Independent) suggesting that the dynamics of PFC are remarkably changed when animals are actively engaged in communicative exchanges relative to other contexts in which the animals hear their vocalizations.

To further explore the Antiphonal context, we investigated neural activity during the reciprocal conversational exchanges that occurred between the subject and the VM. The Antiphonal context was divided into three different categories based on the length of the conversational exchange. we classified single responses by the subject with no subsequent reengagement as “No Conversation.” Conversations in which subjects vocally responded reciprocally two to three times were labelled as “Short Conversation.” Lastly, any conversation that had at least 4 consecutive reciprocal exchanges between subject and VM were “Long Conversation.” Figure 2.9A shows the heatmap of the single events found within these three categories. This set of data includes more units than in Figure 2.8 due to only looking at units that were maintained during the Interactive sessions without regard to the other sessions. In general, there was some difference in mean response across the three conversation lengths. None of the values were reminiscent of Restrained of Freely conditions though. Figure 2.9B plots the ratio of trials labeled as a given conversation length for a single unit over the total trials (Actual Ratio) versus the ratio of significant single events in that conversation over the total trials (Significant Trial Ratio). If there was no bias for the amount of significant events for a given conversation length, then the two ratios would be equal because the probability for the significant events to occur within a particular conversation would be the same as the probability of the conversation occurring. Most sessions would either be on the line or underneath the unity line in such cases. As we see across the three conversation lengths, however, the least-squares line fit across the three lengths progressively increases in slope. Each of the three conversation length groups yielded a significant positive correlation. Furthermore, the ratio of units above the unity line (and thus a higher than expected amount of significant single events) increased with conversation length (None: 31.30%, Short: 45.53%, Long: 64.46%). By looking at the unit data

of significant trial ratio divided by Actual ratio, there was also significant difference found on the mean unit ratio (1-Way ANOVA df = 733, p = 0.011). Figure 2.9C show the No and Short conversation lengths both had confidence intervals across 1 with no significant difference between them. Meanwhile, the Long conversation was significantly higher than No conversation and had 95% confidence intervals above 1 (Tukey-Kramer corrected, No*Long p = 0.009, No conversation: [0.926 to 0.992], Short conversation: [0.973 to 1.039], Long conversation: [1.065 to 1.131]). This indicates that there was a general trend of higher significant trials in the longest conversations above what would be expected. It was also higher than 'No conversations' events which suggests the dynamics of PFC activity during active communication scale based on the length of the reciprocal vocal interaction.

2.4.6 Single Event Driven Response

The preceding analyses indicate that natural communication itself is not a singular context, but a dynamic process that is best conceptualized by a compilation of distinct events that are supported by idiosyncratic neural mechanisms in PFC. To more directly test this question, we leveraged a novel VM behavioral paradigm previously developed in the lab designed to test whether a change in stimulus category membership is meaningful to marmosets during conversational exchanges (Miller and Thomas, 2012). This playback paradigm is similar to the design employed in other parts of the current experiment, but during conversational exchanges a test stimulus is broadcast once the conversation length reaches 2-3 consecutive, reciprocal exchanges (Figure 2.10A). In half the instances, this test stimulus is a Probe; a vocalization that differs from the expected stimulus class. In the other 50% of trials it is a Control; a vocalization that is consistent with the expected stimulus. For these experiments, the Probe stimulus was either a different caller (Identity change) or a reversed phee call (Acoustic change). We only

analyzed units from test sessions that resulted in at least five probe and control trials. Within those units, we compared the overall firing rate during Probe trials and Control trials. Overall, we observed that 31/97 neurons recorded in sessions that reached these behavioral thresholds exhibited a significant difference in activity between Probe and Control Trials. Figure 2.10B shows the mean normalized PSTH for units that had a significant preference for the Probe (blue, 23/97 units, 23.71%) and those with a significant response to Control (red, 8/97 units, 8.25%). The data for each was split into five categories: calls that occurred prior to Probe or Control trials, the Probe or Control itself, and the subsequent call if the subject responds to either one. Units with a preference for Probe trials had a significantly higher response than in Control.Identity change contributed 21 probe-preferring units out of 71 total units (29.6%) while Acoustic change contributed only 2/26 (7.69%).. These data suggest that PFC neurons are acutely sensitive to socially meaningful single events in natural communicative exchanges, such as those that deviate from expectations.

2.4.7 Latency to Peak Response

In the final analysis, we compared the latency to peak response across Phee presented in six different contexts in these experiments. Specifically, we examined neural response from the population of neurons with stable recordings across the Interactive, Restrained, and Freely that exhibited a significant change in firing rate for Phee stimuli within a given context, including distinguishing between the Antiphonal and Independent events in the Interactive context. Finally, we included Probe preferring responsive units described above. Figure 2.10C plots the mean normalized PSTH and 95% CI for each of these six contexts. As is evident in this figure, the Probe had a notably longer latency to peak response compared to the other mean normalized PSTH values (750 msec compared to 450 msec for Restrained and Freely, 550 msec for

Independent calls). This suggests that identifying a meaningful category change within a conversation may rely on distinct mechanisms in PFC that necessitate additional processing.

2.5 Discussion.

Here we sought to test the contextual effects on primate prefrontal cortex neurons in marmoset monkeys by comparing within-neuron responses to acoustic stimuli in a series of conditions. This tactic was designed to ascertain whether neural responses using more traditional approaches to studies of the primate brain are predictive of the presumptive natural analog for the first time. Results consistently indicated that vocal signal processing in PFC neurons was substantially affected by context. The same neurons that exhibited robust stimulus driven responses to vocalization stimuli when animals were head-restrained, typically exhibited a significantly weaker or no response to the exact same vocalization stimuli when subjects were engaged in natural vocal interactions. Furthermore, these differences were not simply due to differences in mobility, nor as a result of the neurons spatial receptive fields. Rather, these data indicate that at least for prefrontal cortex neurons, vocal signal processing in the traditional paradigms routinely used (Gifford et al., 2005; Romanski et al., 2005; Averbeck and Romanski, 2006; Romanski and Averbeck, 2009; Plakke et al., 2013a) are not predictive of the same process during natural communication. Because these results are consistent with findings in a recent study of face cells (McMahon et al., 2015), the influence of context on social signal processing may be far more profound than is typically considered.

These experiments first compared neural responses across two contexts – Restrained & Freely-Moving. By presenting subjects with the identical acoustic stimuli and presentation pattern in the Restrained and Freely-Moving contexts, we sought to directly test whether subjects' mobility affected neural activity. We observed that individual neurons recorded stably

across both contexts typically exhibited relatively similar responses to noise or vocalization stimuli. In other words, units exhibiting a significant response to noise or vocalizations in one context was typically also responsive to the same stimulus in the other context. Context did, however, affect some modest changes in the properties of the neural response. Specifically, these units did exhibit a moderately lower firing rate in Freely-Moving than in Restrained context (Figure 2.3A & B). Furthermore, these units also exhibited a greater variance in the neural response in the Freely-Moving context (Figure 2.3D). These data suggest that the mobility of the animal had little effect on whether a single neuron was responsive to an acoustic stimulus, though it did affect some moderate properties of the neural response itself.

To further explicate the contextual effects on vocal signal processing, we next examined neurons recorded across Restrained, Freely-Moving and Interactive. The interactive context involved directly engaging subjects in their natural vocal interactions to ascertain how neural responses to phee calls were affected by the dynamics of natural communication. When characterizing within-unit differences, we observed broad contextual affects across the population (Figure 2.4A). While many neurons exhibited significant changes in activity across only a single context, a pattern did emerge of pulse-based response within the phees (Figure 2.4B). The most robust responses were for neurons responsive to phee calls in both the Restrained and Freely contexts (Figure 2.5A & Figure 2.6A), or across all three contexts (Figure 2.5B), though in these latter units the response in the Interactive context was notably more modest than in the other two (Figure 2.6B). As it turned out, the Independent calls produced by the VM elicited similar responses to the passive-listening contexts, while the actual engagement with the VM in conversations elicited a more tepid response (Figure 2.6C). Given that this is when active communication occurs, these results are consistent with our prior work showing

limited vocalization responsive frontal cortex neurons during active conversations (Miller et al., 2015). More broadly, these results suggest that while PFC responses during Freely and Restrained are remarkably consistent, that the context of natural communication relies on distinct neural mechanisms leading to different responses. Actively engaging in communication exchanges is likely supported by a myriad of processes related to the dynamic nature of the behavior itself beyond simply representing the sound. These differences are further evident in the subsequent analyses.

The difference in response to context led us to look at the normalized firing rate across the maintained units within the three contexts. We found a significant difference in the firing rate across all trials of Interactive compared to Restrained and Freely (Figure 2.7A). This distribution difference was further explored by attempts to classify trials by contexts across populations of subsets of maintained units (Figure 2.7B). We found significant performance increase by including any responsive unit or all units compared to those that had significant response in all three contexts. Showing similarity to the distribution differences, the classifiers performed the worst at distinguishing between Restrained and Freely trials compared to Interactive (Figure 2.7C). We then tested the difference in distributions between Antiphonal and Independent events within the Interactive Contexts and found no significant difference (Figure 2.7D). Classifiers performed significantly higher when we included All Units rather than just ones that had a preference for any context (Figure 2.7E). We also found that the Antiphonal class was more likely to be miscategorized as Independent than the reverse (Figure 2.7F), suggesting that there were some trials that had similarly robust responses as seen in Independent.

Analyses indicated that considering the Interactive context as a singular process was too coarse because of the dynamic nature of natural communication. As a result, we sought to focus

analyses on single events, rather than averaging responses across a test session, as a way of guiding our understanding of the role PFC plays in representing vocalizations. In these analyses, we deemed a significant event to be each instance a phee stimulus was broadcast and elicited a change in firing rate at least 2SD above baseline. While most units had at least one significant event a session, the average responses in each context recapitulated our prior results (Figure 2.8A). What was interesting to note was the change in the ratio of significant trials (Figure 2.8B) and the amount of time units spent above threshold (Figure 2.8C). In both cases, the Interactive context performed the worst suggesting that this context affected the least change in PFC activity. This was further emphasized when splitting the Interactive context once more between the Antiphonal and Independent calls (Figure 2.8D). While both had tepid overall responses in comparison to Restrained and Freely, they each had similar rates for significant events (Figure 2.8E). But, once more, the Antiphonal condition was significantly different from Independent when looking at the rate the significant events spent above the threshold (Figure 2.8F). An important event within the Interactive context, however, are the natural conversations that emerge. To explicate whether these distinct behaviors offered further insight, we further analyzed PFC activity at the single event level based on the length of conversations (Figure 2.9A). Results indicated some modest changes by conversation length, but by analyzing the ratio of significant single events to the overall ratio of those various categories (Figure 2.9B), we found that longest conversations had significantly higher ratio of significant events than expected (Figure 2.9C). These data suggest that while PFC is often only moderately responsive when hearing phees during active communication, the conversations themselves are a distinct event that drive PFC mechanisms.

A notable shortcoming to these data is that we did not compare neural activity directly between analogous conditioned and natural behaviors. However, neural responses to vocalizations in macaque prefrontal cortex when animals are trained to perform behavioral tasks involving vocalization stimuli exhibit patterns of activity that are notably different than what we observed here (Cohen et al., 2009; Plakke et al., 2013b; Hwang and Romanski, 2015). Cohen and colleagues (Cohen et al., 2009) for example, recorded vlPFC neurons while macaque monkeys performed a category detection task involving multiple vocalization types. While many neurons in the population exhibited strong responses related to the specific task demand, others exhibited strongly driven responses to the vocalization stimuli themselves irrespective of the task. By contrast, we observed more modest neural modulations during natural communication here, even during active conversational exchanges. Perhaps the more direct parallel, however, would be the pattern of responses observed during the Probe condition performed in our experiments (Figure 2.10A). Like the task employed by Cohen et al (2009) a population of neurons exhibited strongly driven activity only when a change in category was detected (Figure 2.10B). Though here the category was true only for a change in caller identity, rather than call type, as was reported in the earlier study (Cohen et al., 2009). Notably, we also found that the normalized PSTH had a different peak response with a delay of 300 msec compared to other similar presentations of Phee calls in different contexts (Figure 2.10C), suggesting extra cognitive load in expectation violation. Overall, the Probe contexts suggests that our data are consistent with the notion that PFC plays a crucial role in identifying important changes in the world, but that under natural circumstances this may only manifest for isolated, particularly important single events. Whereas in other natural contexts, even when animals are engaged in coordinated social interactions that

require attention to a conspecifics behavior to coordinate their own behavior, different mechanisms occur.

These data suggest that explicating the neurobiology of social communication in the primate brain likely necessitates studying the brain while animals engage in these dynamic, natural behaviors. Reductionistic approaches typical of neuroscience are not without their limitations, but precisely what these limitations may be is rarely considered. Conceptions that stimulus presentations in the absence of behavior affords insight into the core, foundational organization of the brain, upon which mechanisms to support behaviors simply sum additively is not likely to be strictly true, at least for social communication. The novel experiment described here was the first to directly test how the context of more traditional primate neuroscience paradigms affects neural responses relative to the natural analog, and even the single communicative behavior here may not be sufficient to fully appreciate how context affects neuronal processes. Behavior, after all, is not a singular monolith. Species possess highly diverse behavioral repertoires. For primates, the corpus of behaviors within the primate social domain is the most distinguishing characteristic of the Order (Miller et al., 2016). The emergence of computer vision, machine-learning methods for quantifying behaviors through video analysis, along with wireless neural recording systems, offers exciting opportunities to investigate and model the complexities of the primate social brain in a way not previously possible (Calhoun et al., 2019). Such a computational neuroethological approach offers exciting opportunities to untether our conceptual and quantitative understanding of primate brain function.

2.6 Acknowledgements

This work supported by the National Institutes of Health Grant R01 DC012087 to C.T.M.

2.7 Figures

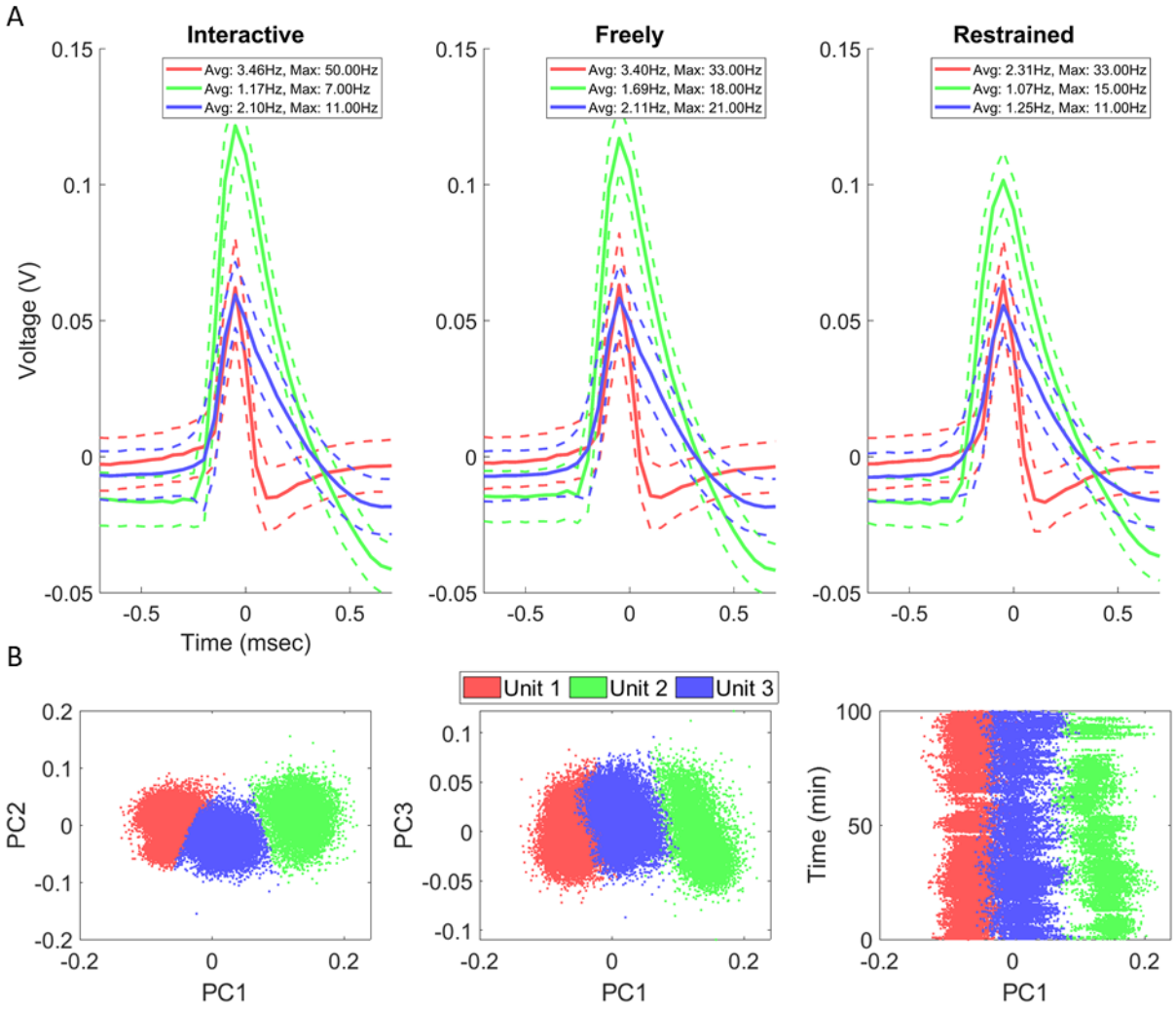


Figure 2.1: Example of single unit stability across contexts within a single recording session. (A) Waveform stability for three single units across the three-contexts recording session. Solid lines represent the mean waveform for a unit with dashed lines representing the standard deviation above and below the mean. (B) PCA space of all waveforms with each unit clustered by color across all three contexts. Units colored by k-means clustering to verify the thresholding method used. Units maintained their shape across time given the three columns seen in the last graph.

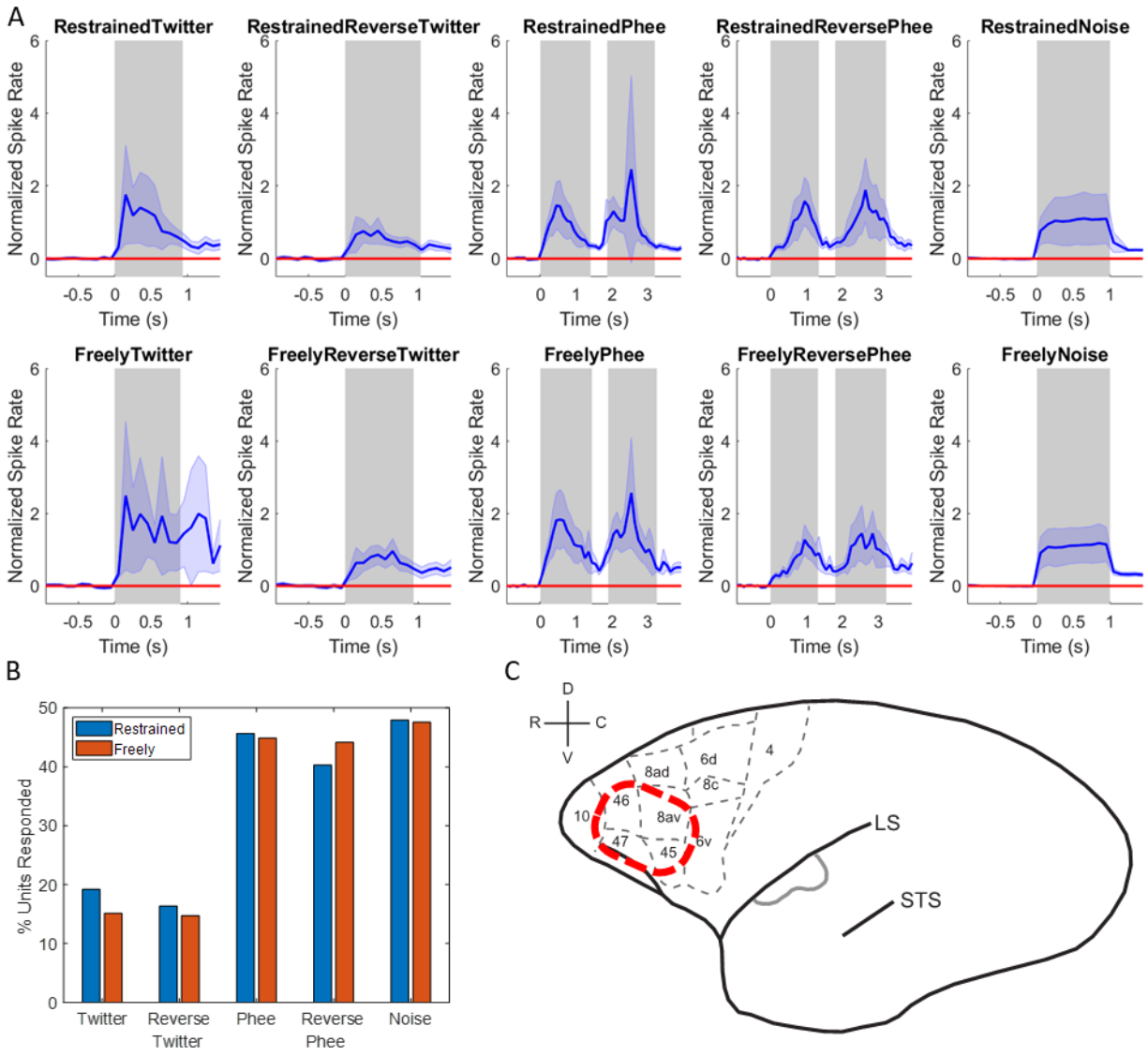
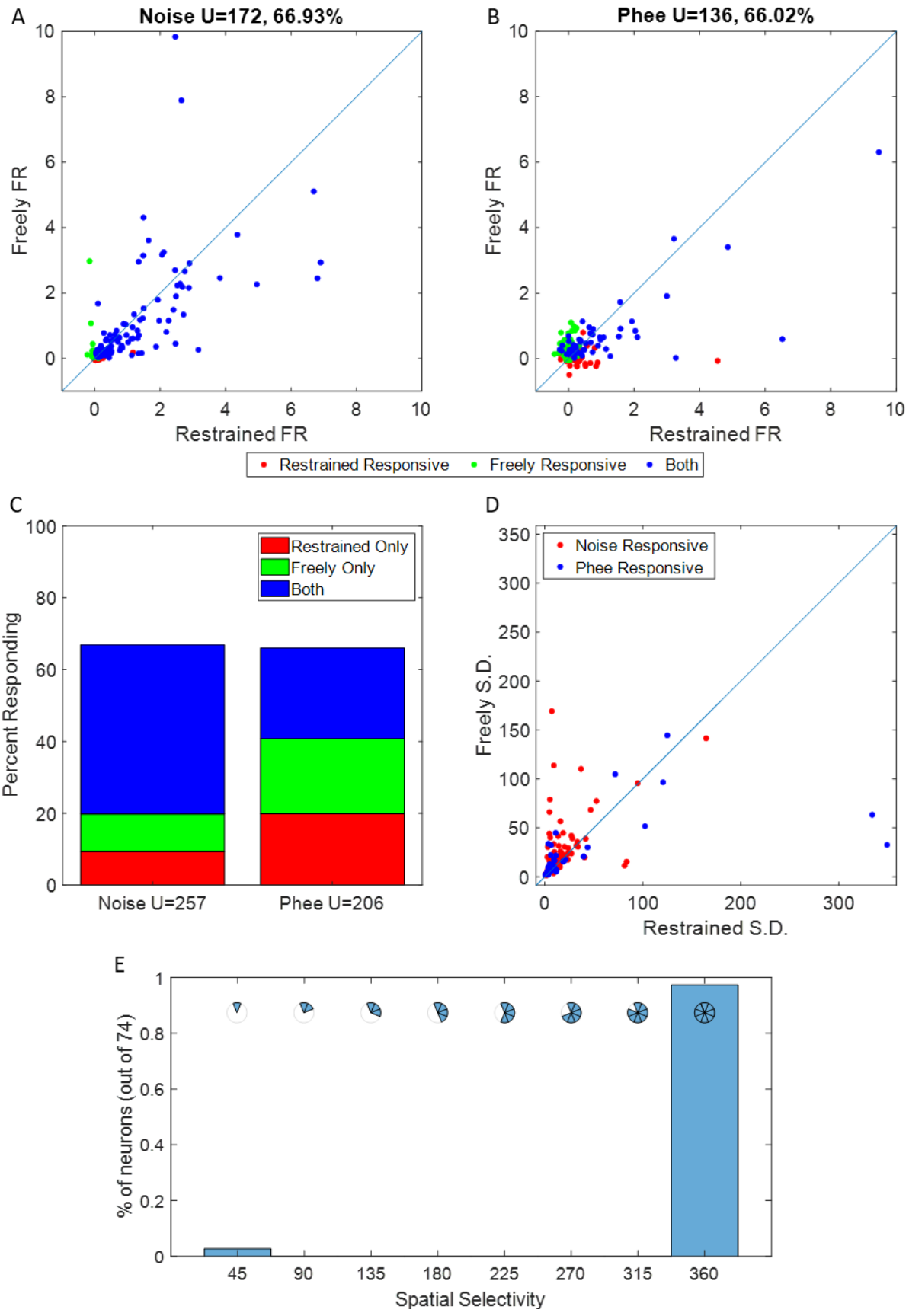


Figure 2.2: Responsiveness of all units across Restrained and Freely contexts. (A) Normalized PSTH for all units that had a significant response to a titled stimulus set in the Restrained and Freely contexts. Any unit with a response was included. The solid blue line is the mean normalized PSTH for all the units involved. Grey bars represent the average duration of the stimulus. Phees had two pulse-calls and thus two bars. Red line represents the mean firing rate prior to onset. Shaded blue area represents the confidence interval. (B) Overall responsiveness per stimulus category and context. Any unit found within that context that was exposed to the particular stimulus set is included. The percentages represent the number of those units that had a significant response for the given stimulus in that context. (C) Anatomical map of the frontal cortex of the common marmoset. The dashed red outline represents roughly the area that was explored with the arrays.

Figure 2.3: Effect of mobility on Firing Rate and variance. (A & B) Comparison of Noise and Phee responding units that were maintained across the two contexts (Restrained and Freely). Only units that had at least a response in one context is shown. Non-responding units are excluded. Titles of the graphs show how many units had at least one response. FR refers to the normalized Firing Rate for each unit in comparison to prior onset (outlined further in Methods). Blue line represents the unity line. (C) Stacked bar graph showing percentage of maintained units broken down by their response between Restrained only, Freely only, and Both. (D) Comparison of standard deviation for the normalized Firing Rates between Restrained and Freely contexts for Noise and Phee responsive units. Only the units that had a significant response in both contexts was included. Blue line represents the unity line. (E) Distribution of units found to have any selectivity for Noise in the Orientation context. Only two units had selectivity for one angle of eight. The pie charts above represent the number of degrees that a unit had orientation tuning for. One shaded section means only one orientation was significant. All eight shaded means the unit had general significant response to Noise but not any one orientation (71 units), or significant response to all eight orientation (1 unit).


A
Noise U=172, 66.93%
Freely FR
Restrained FR
B
Phee U=136, 66.02%
Freely FR
Restrained FR
Restrained Responsive
Freely Responsive
Both
C
Percent Responding
Restrained Only
Freely Only
Both
Noise U=257
Phee U=206
D
Noise Responsive
Phee Responsive
Freely S.D.
Restrained S.D.
E
% of neurons (out of 74)
Spatial Selectivity

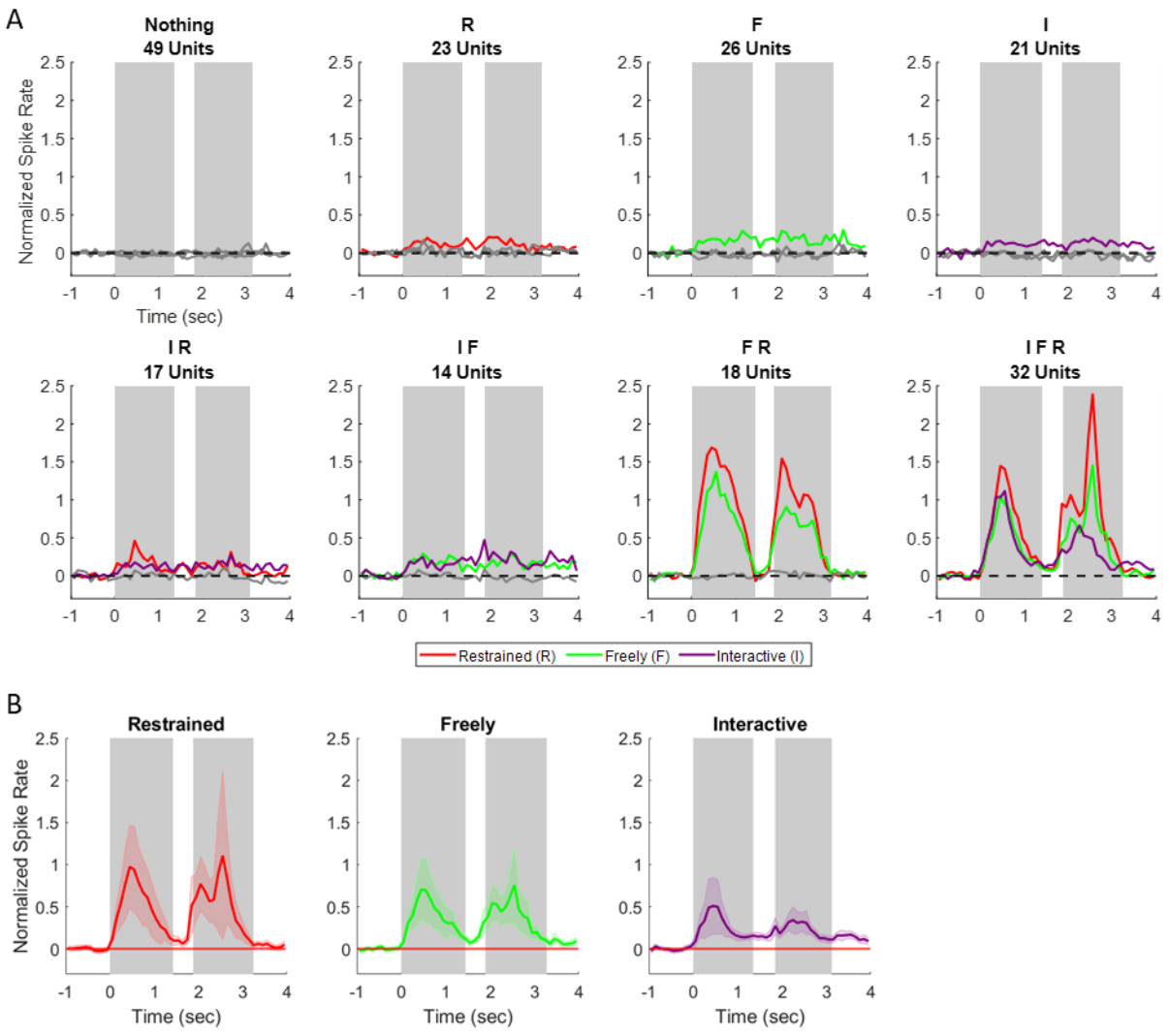
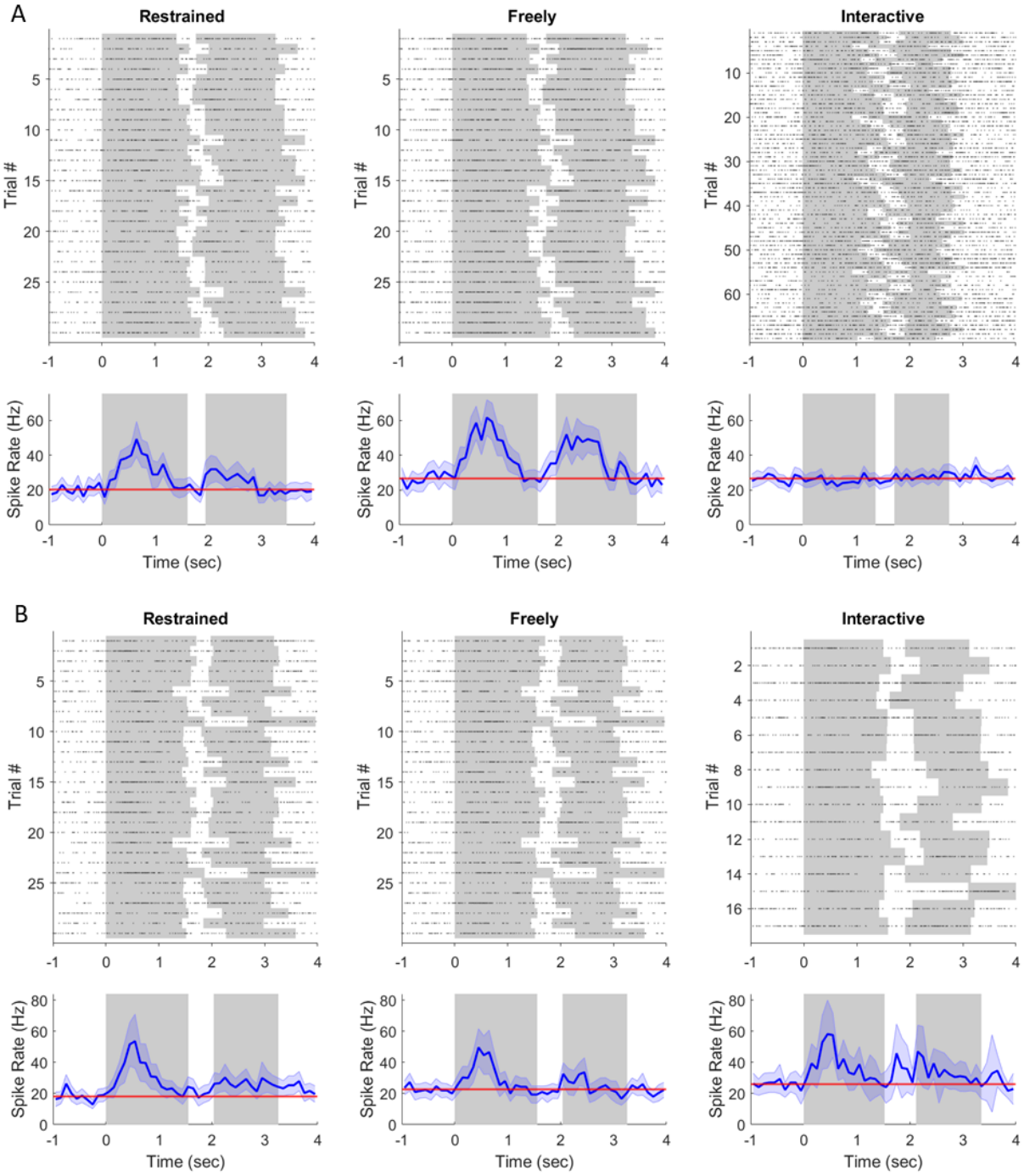


Figure 2.4: Maintained units for Restrained, Freely, and Interactive and their response categories to phee stimuli. (A) Mean normalized PSTH for each set of units found to have significant response to phee calls in the given context listed in the title. For three contexts, there are eight possible combinations of significant response. Overall, 200 units were maintained with 151 having a response in at least one context. Grey bars represent the average pulse duration for the units in that category. Dashed black line represents the mean firing rate prior to onset of stimulus. Colored solid lines represent the mean normalized PSTH for units in that category. Grey color represents the non-responsive units. The red, green and purple lines represent the Restrained, Freely, and Interactive contexts that a unit had significant response for. (B) Mean normalized PSTH for all units that had any response for each context: Restrained, Freely, and Interactive. Colored lines represent the mean normalized PSTH for all the units with 95% confidence intervals as the shading. Grey bars represent the mean duration of each pulse for all calls presented to these units.

Figure 2.5: Exemplar units maintained across Restrained, Freely, and Interactive contexts. (A) Exemplar unit showing a significant response for Restrained and, Freely but not Interactive phee calls. The top part for each column is a raster plot. Each dot represents a spike for the given unit. Grey bars represent the duration of the stimulus; all of which have 2 pulses or bars. Below each raster is the average raw firing rate for the given unit. The grey bars represent the mean duration of each gray bar, all aligned at onset. The red line represents the mean firing rate prior to onset. Solid blue line represents the mean firing rate, and the shaded area represents the 95% Confidence Interval. (B) Exemplar unit showing a response for the Restrained, Freely, and Interactive. This unit had fewer calls for the Interactive condition as the subject responded less frequently and was less engaged than the recording session for the subject in (A).


A
Restrained
Freely
Interactive
Trial #
Spike Rate (Hz)
Time (sec)
B
Restrained
Freely
Interactive
Trial #
Spike Rate (Hz)
Time (sec)

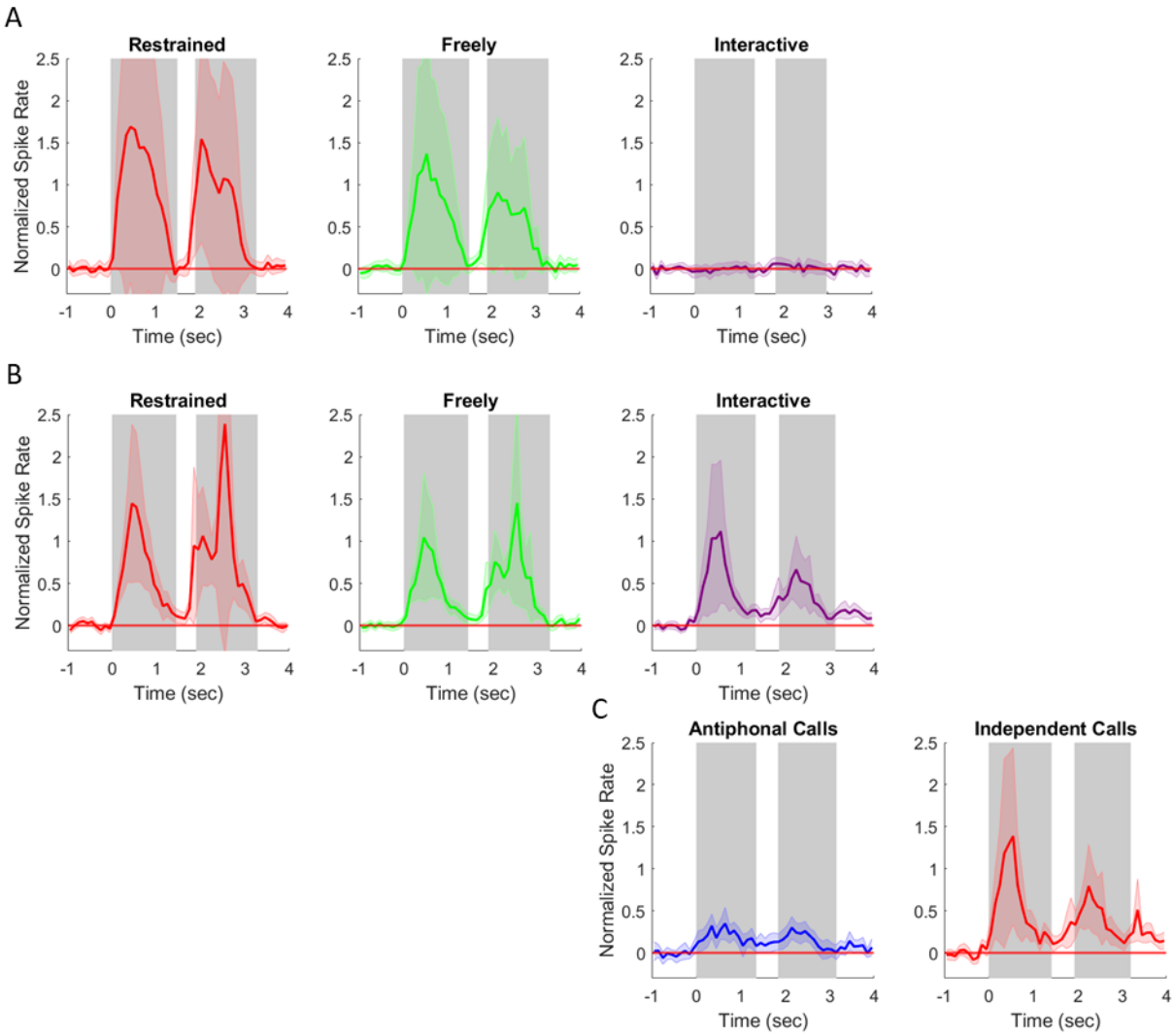


Figure 2.6: Maintained unit responsiveness in FR and IFR. (A) Mean normalized PSTH for all units that had significant response to Phees for in FR (Freely and Restrained). (B) Mean normalized PSTH for all units that had significant response to Phees for all three contexts (Interactive, Freely, and Restrained). (C) The Interactive context was split between calls the VM made in response to the subject (Antiphonal Calls) and calls made without a prior response (Independent Calls). Grey bars represent the mean duration of each pulse for the two-pulse phee calls played by the VM. Solid blue line represents the mean normalized PSTH for all 32 units for that context. Shaded areas represent the 95% confidence interval.

Figure 2.7: Classification results of predicting Restrained, Freely, and Interactive, and Antiphonal and Independent Calls. (A) Distribution of normalizing Firing Rates for all units that had a response to Phee calls in all three contexts (32 Units, Figure 2.5A). All trials for each unit was combined to plot the distributions on Normal Probability plots which compares the distribution of the actual data (blue crosses) to the hypothetical normal distribution it should come from (dashed red line). Each distribution indicates a significant difference from a normal distribution and a right skew of the actual distribution. Interactive distribution was significantly different from Restrained and Freely. (B) Box plots of the results of 1000 simulations for classification of Restrained, Freely, and Interactive trials, across three different data sets. IFR refers to the 32 units in (A). Any Response refers to the 151 units with any kind of response (Figure 2.5B). All Units refers to using data from all 200 units. MCC refers to the Matthews Correlation Coefficient which gives a value to the performance of correctly predicting the classes from training to test data. 0 refers to random chance, and anything below it is worse than guessing, with +1 representing perfect classification of test data. IFR was significantly worse in performance compared to Any Response and All Units. (C) Mean confusion matrix for each of the data sets and their 1000 simulations. Horizontal rows are normalized to each other and show what percentage each class was predicted to be. Bottom two rows represent the overall performance of prediction for each class. Top row is the correct percentage, and bottom is the incorrect. (D) Distribution of normalizing Firing Rates for all units that had a response to Phee calls in all three contexts and only their Antiphonal Calls and Spontaneous calls from the Interactive Context (32 Units, Figure 2.5C). Distributions were not normal and skewed right with no significant difference between them. (E) Classification performance of the three data sets to classify between Antiphonal Calls and Independent calls within Interactive context. Each data set was significantly different from the other two with IFR at significantly lower response than Any Response and All Units. All Units had significantly higher response than both. (F) Mean confusion matrix for Antiphonal class versus Independent class. Significant difference of a given data set's performance from the other two is represented by an asterisk ($p < 0.001$).

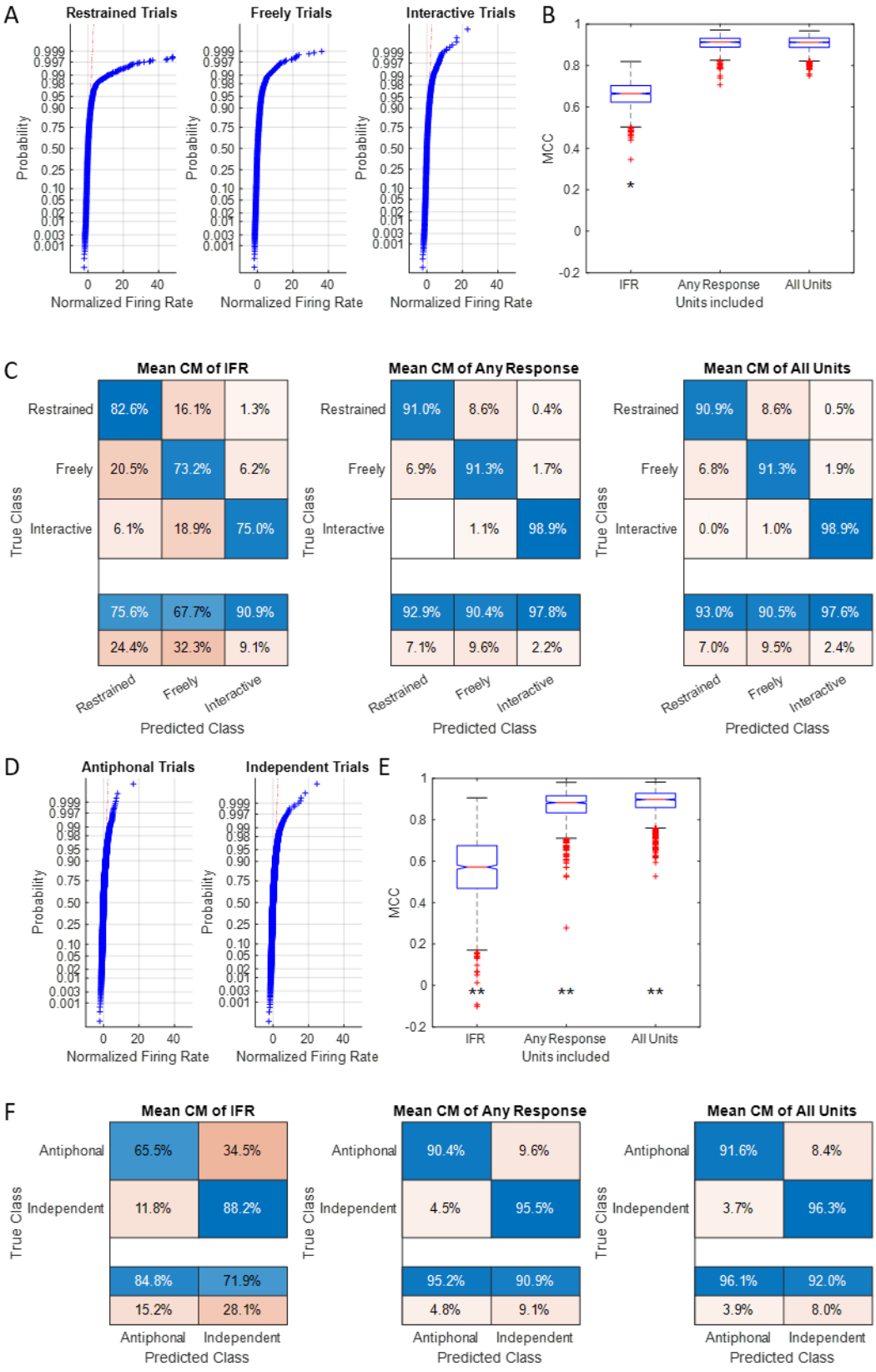

A
Restrained Trials
Freely Trials
Interactive Trials
Probability
Normalized Firing Rate
B
MCC
IFR
Any Response
All Units
Units included
C
Mean CM of IFR
Mean CM of Any Response
Mean CM of All Units
True Class
Predicted Class
Restrained
Freely
Interactive
82.6% 16.1% 1.3%
20.5% 73.2% 6.2%
6.1% 18.9% 75.0%
75.6% 67.7% 90.9%
24.4% 32.3% 9.1%
91.0% 8.6% 0.4%
6.9% 91.3% 1.7%
1.1% 98.9%
92.9% 90.4% 97.8%
7.1% 9.6% 2.2%
90.9% 8.6% 0.5%
6.8% 91.3% 1.9%
0.0% 1.0% 98.9%
93.0% 90.5% 97.6%
7.0% 9.5% 2.4%
D
Antiphonal Trials
Independent Trials
E
F
Mean CM of IFR
Mean CM of Any Response
Mean CM of All Units
Antiphonal
Independent
65.5% 34.5%
11.8% 88.2%
84.8% 71.9%
15.2% 28.1%
90.4% 9.6%
4.5% 95.5%
95.2% 90.9%
4.8% 9.1%
91.6% 8.4%
3.7% 96.3%
96.1% 92.0%
3.9% 8.0%

Figure 2.8: Significant single events across the units maintained in all any context.Heatmap of all significant trials for any unit that was held in the given column's context. Each trial is represented as a row on the heat map with 100 interpolated points from 300 msec prior to onset of the call and 4300 msec post. Colors range from cyan to yellow representing points that range in the Z-Score from -2 to 2. Anything above those ranges is capped at the max values. Trials were sorted by position of first value greater than 2 SD. The plots below each heatmap represent the mean values for each of the columns with a 95% confidence interval shading in blue. Grey lines represent the borders of the onset and offset of the average phee call within these trials. (B) The mean and confidence intervals of the ratio of significant trials to total trials for each unit included in (A). Interactive was significantly lower than Restrained or Freely conditions. (C) The mean and confidence intervals of the average ratio of time spent above a threshold for each of significant trials. Interactive was significantly lower than Freely and Restrained contexts. (D) Heatmap of all significant trials for all units that were maintained across the three contexts and had at least 5 Antiphonal call trials and 5 Independent call trials. (E) Ratio of significant trials to all trials for all the units included in each context heat map. (F) Ratio of time each significant single event was at or above the 2 SD threshold. Significant difference of a given set to all other sets is represented by an asterisk.

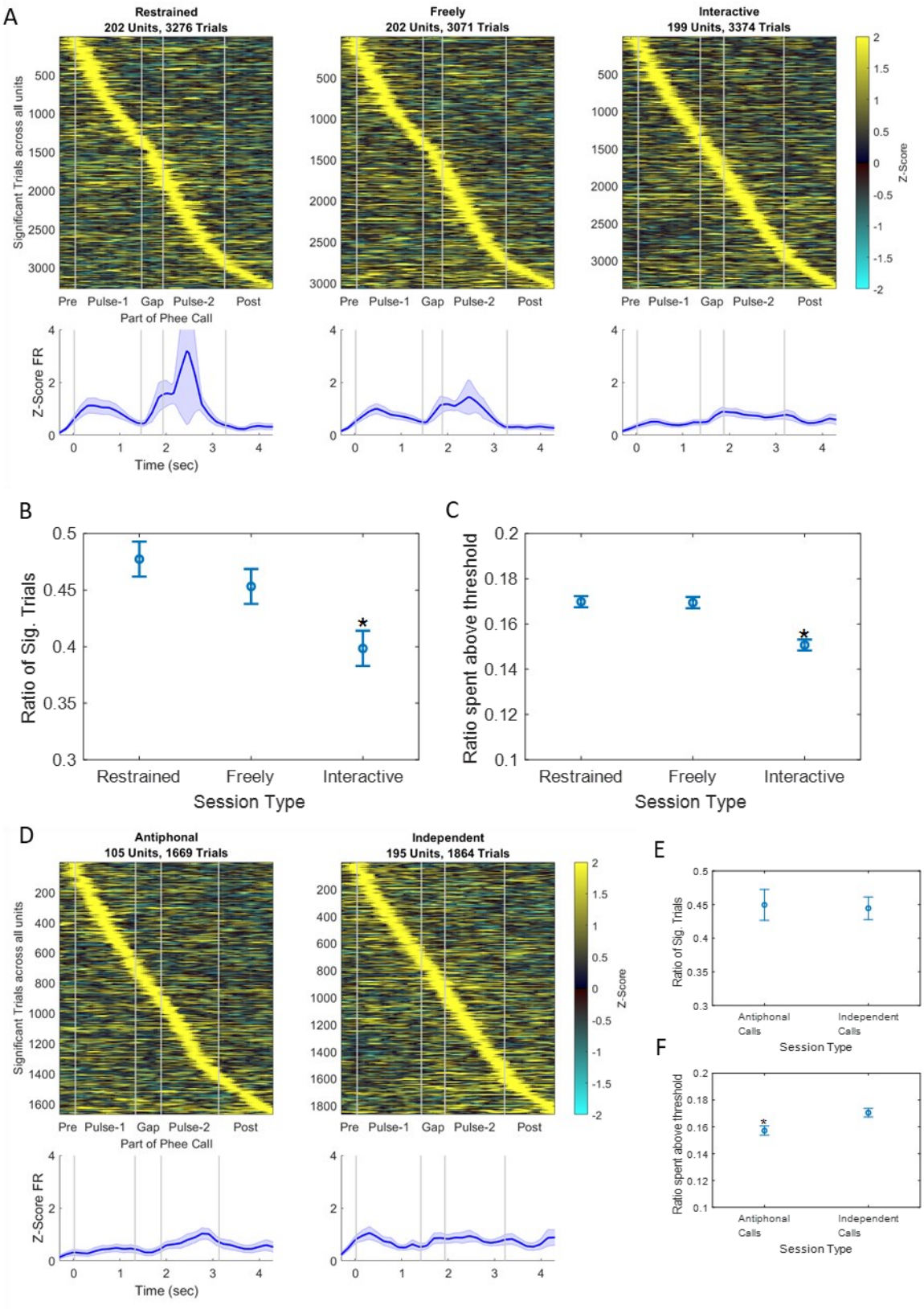

A
Restrained
202 Units, 3276 Trials
Freely
202 Units, 3071 Trials
Interactive
199 Units, 3374 Trials
Significant Trials across all units
Pre Pulse-1 Gap Pulse-2 Post
Part of Phee Call
Z-Score
Z-Score FR
Time (sec)
B
Ratio of Sig. Trials
Restrained Freely Interactive
Session Type
C
Ratio spent above threshold
Restrained Freely Interactive
Session Type
D
Antiphonal
105 Units, 1669 Trials
Independent
195 Units, 1864 Trials
Significant Trials across all units
Pre Pulse-1 Gap Pulse-2 Post
Part of Phee Call
Z-Score
Z-Score FR
Time (sec)
E
Ratio of Sig. Trials
Antiphonal Calls
Independent Calls
Session Type
F
Ratio spent above threshold
Antiphonal Calls
Independent Calls
Session Type

Figure 2.9: Significant single events across the units that were found in Interactive context grouped by No Response, Short conversation, and Long Conversation. (A) Heatmap of all significant trials for any unit that was held in for the three conversation lengths. Each trial is represented as a row on the heat map with 100 interpolated points from 300 msec prior to onset of the call and 4300 msec post. Colors range from cyan to yellow representing points that range in the Z-Score from -2 to 2. Anything above those ranges is capped at the max values. Trials were sorted by position of first value greater than 2 SD. The plots below each heatmap represent the mean values for each of the columns with a 95% confidence interval shading in blue. Grey lines represent the borders of the onset and offset of the average phee call within these trials. (B) Scatter plot for each conversation length to show whether a given unit has a higher than expected amount of significant events. The x-axis is the ratio of trials for that unit that were labeled as part of each conversation length. The y-axis is the ratio of significant trials that were of the given conversation length in comparison to all trials. The dashed black line represents unity and anything above the line has a higher than expected amount of significant trials for that conversation length. Blue dots represent each unit used. Blue line represents the fitted line through the unit data. (C) The comparison of ratio of significant trial ratio to actual trial ratio (y-axis over x-axis) for each unit across conversation lengths. Long conversations were significantly higher than No Response. Error bars represent 95% confidence intervals. Significant difference of a set compared to another is signified with an asterisk.

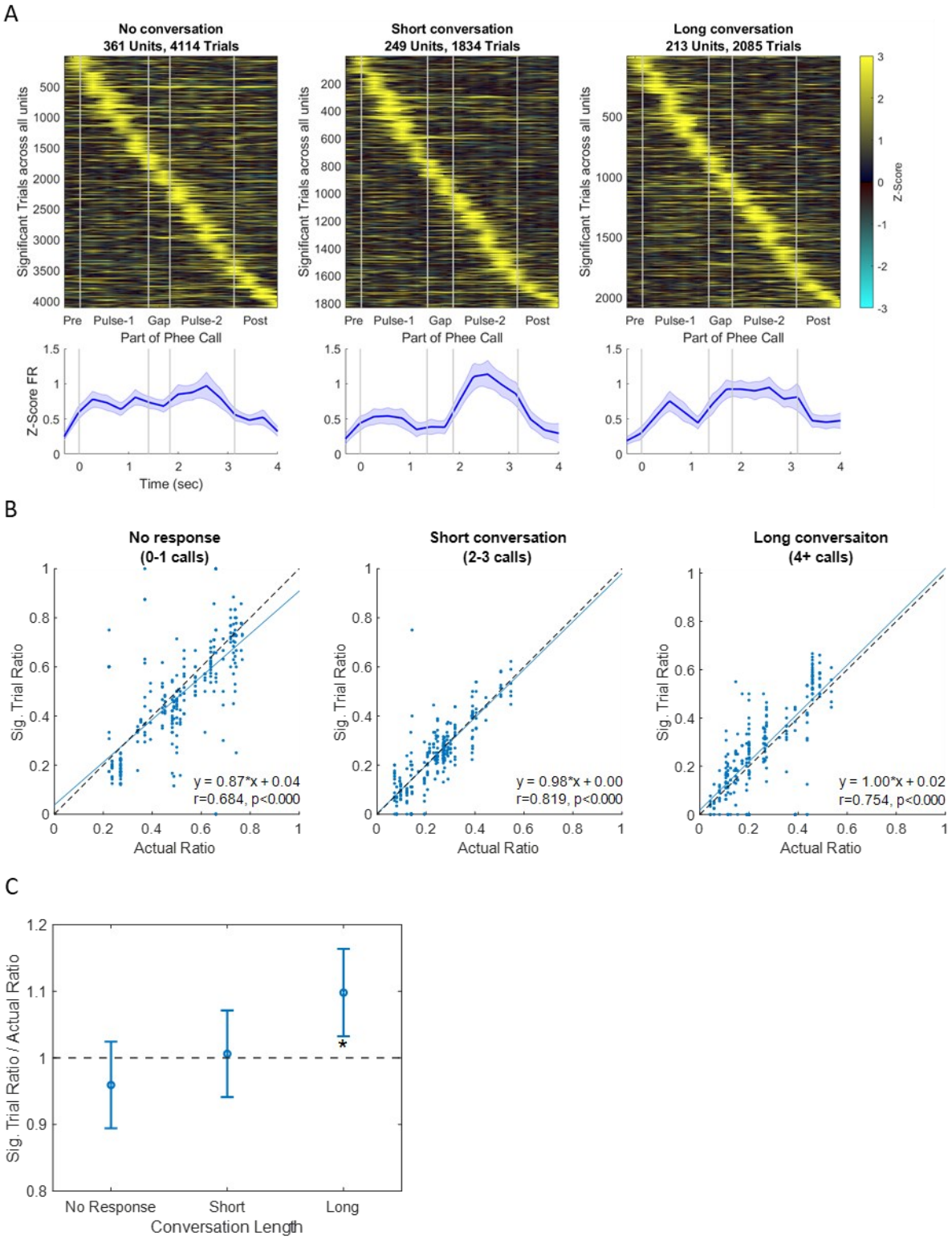
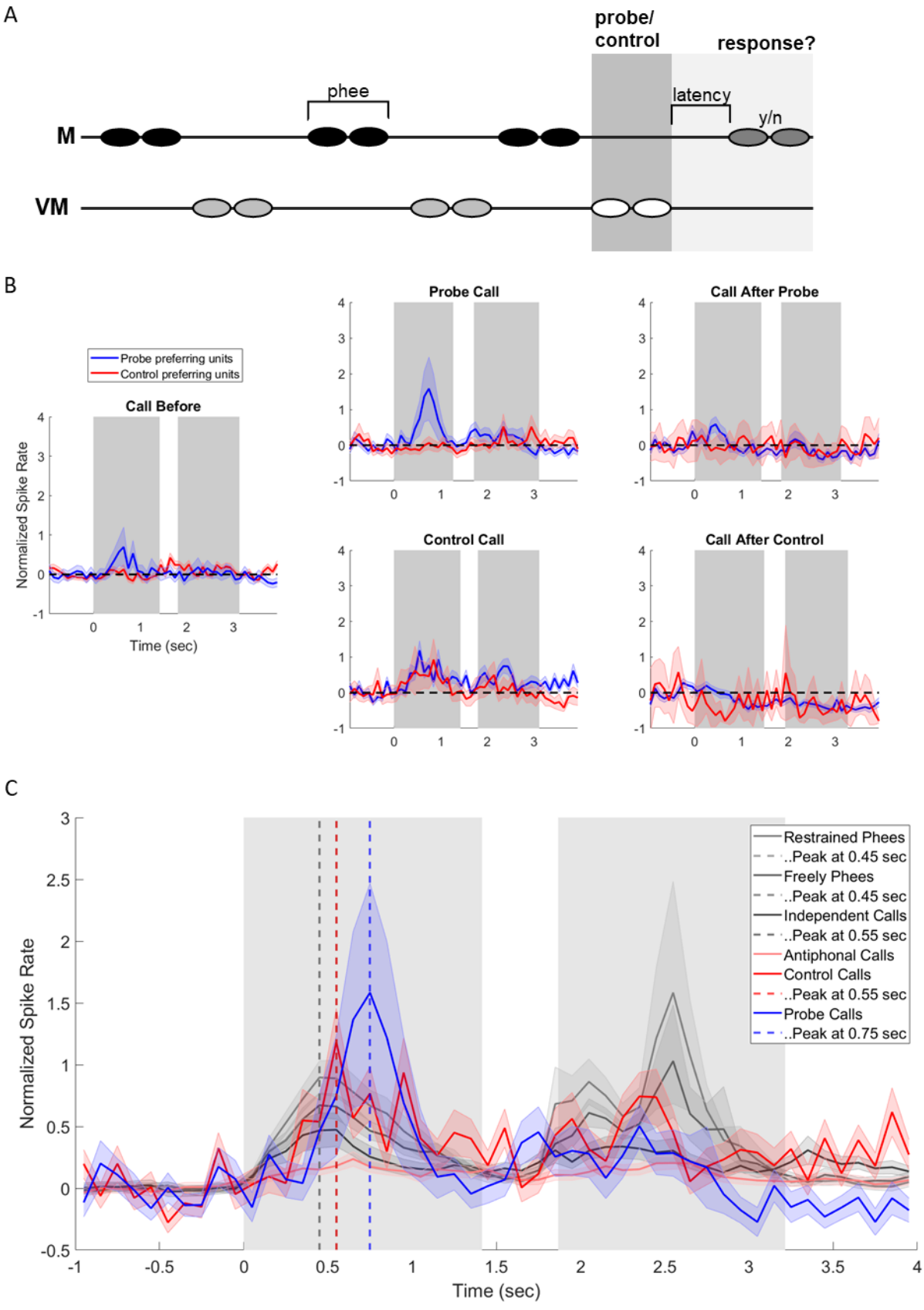

A
No conversation
361 Units, 4114 Trials
Short conversation
249 Units, 1834 Trials
Long conversation
213 Units, 2085 Trials
Significant Trials across all units
Pre
Pulse-1
Gap
Pulse-2
Post
Part of Phee Call
Z-Score
Z-Score FR
Time (sec)
B
No response
(0-1 calls)
Short conversation
(2-3 calls)
Long conversaiton
(4+ calls)
Sig. Trial Ratio
Actual Ratio
y = 0.87*x + 0.04
r=0.684, p<0.000
y = 0.98*x + 0.00
r=0.819, p<0.000
y = 1.00*x + 0.02
r=0.754, p<0.000
C
Sig. Trial Ratio / Actual Ratio
No Response
Short
Long
Conversation Length

Figure 2.10: Driving single event response with Probe paradigm. (A) Outline of the Probe paradigm. Subjects engage in two to three conversation exchanges with the VM. On the second or third VM call, there is a 50% chance that VM will play a phee call from the expected caller (Control). The other 50% means a phee call from another previously recorded caller. Ovals represent the pulses of a phee call which are typically two pulses. (B) Mean normalized PSTH for Probe preferring and Control preferring units. Each set of calls was normalized to 1000 msec prior to onset of the phee calls in that set for each unit. Colored lines represent the mean normalized PSTH. Shaded areas represent their 95% confidence interval. The black dashed line represent the average firing rate prior to onset. Grey bars represent the mean durations of the pulses for each data set. Calls were split into those that were probe and control, the calls prior to each type (combined together), and the immediate calls played after by the VM if the subject responds (separated). Preferring units had a significantly higher response to either probe or control compared to the other. (C) Normalized PSTH plots for Phee responses across the three contexts, the subset of Antiphonal calls, Control calls, and Probe calls. Only units that were Probe preferring were included for Control and Probe. Each solid line represents the mean normalized PSTH for that data set. Shaded areas represent 95% confidence interval. Dotted lines of the same color as data set represent the peak for that normalized PSTH curve. Grey bars represent the mean duration of the pulses of calls for all of the data sets.


A
probe/
control
response?
phee
latency
y/n
M
VM
B
Probe preferring units
Control preferring units
Call Before
Probe Call
Call After Probe
Control Call
Call After Control
Normalized Spike Rate
Time (sec)
C
Restrained Phees
..Peak at 0.45 sec
Freely Phees
..Peak at 0.45 sec
Independent Calls
..Peak at 0.55 sec
Antiphonal Calls
Control Calls
..Peak at 0.55 sec
Probe Calls
..Peak at 0.75 sec
Normalized Spike Rate
Time (sec)

2.8 References

Averbeck BB, Romanski LM (2006) Probabalistic encoding of vocalizations in macaque ventral lateral prefrontal cortex. J Neurosci 26:11023-11033.

Briscoe SD, Ragsdale CW (2018) Homology, neocortex, and the evolution of developmental mechanisms. Science 362:190.

Calhoun AJ, Pillow JW, Murthy M (2019) Unsupervised identification of the internal states that shape natural behavior. Nature Neuroscience 22:2040-2049.

Cohen YE, Russ BE, Davis SJ, Baker AE, Ackelson AL, Nitecki R (2009) A functional role for the ventrolateral prefrontal cortex in non-spatial auditory cognition. PNAS 106:20045.

Freiwald W, Duchaine B, Yovel G (2016) Face Processing Systems: From Neurons to Real-World Social Perception. Ann Rev Neurosci 39:325-346.

Fuster JM (2008) The Prefrontal Cortex. New York: Academic Press.

Gifford GW, MacLean KA, Hauser MD, Cohen YE (2005) The neurophysiology of functionally meaningful categories: macaque ventrolateral prefrontal cortex plays a critical role in spontaneous categorization of species-specific vocalizations. J Cog Neurosci 17:1471-1482.

Guilford T, Dawkins MS (1991) Receiver psychology and the evolution of animal signals. Anim Behav 42:1-14.

Hauser MD (1996) The Evolution of Communication. Cambridge: MIT Press.

Hwang J, Romanski LM (2015) Prefrontal Neuronal Responses during Audiovisual Mnemonic Processing. The Journal of Neuroscience 35:960.

McMahon DB, Russ BE, Elnaiem HD, Kurnikova AI, Leopold DA (2015) Single-Unit Activity during Natural Vision: Diversity, Consistency and Spatial Sensitivity among AF Face Patch Neurons. J Neurosci 35:5537-5548.

Miller CT, Thomas AW (2012) Individual recognition during bouts of antiphonal calling in common marmosets. Journal of Comparative Physiology A 198:337-346.

Miller CT, Beck K, Meade B, Wang X (2009) Antiphonal call timing in marmosets is behaviorally significant: Interactive playback experiments. . Journal of Comparative Physiology A 195:783-789.

Miller CT, Thomas AW, Nummela S, de la Mothe LA (2015) Responses of primate frontal cortex neurons during natural vocal communication. J Neurophys 114:1158-1171.

Miller CT, Hale ME, Okano H, Okabe S, Mitra P (2019) Comparative Principles for Next-Generation Neuroscience. Frontiers in Behavioral Neuroscience 13.

Miller CT, Freiwald W, Leopold DA, Mitchell JF, Silva AC, Wang X (2016) Marmosets: A Neuroscientific Model of Human Social Behavior. Neuron 90:219-233.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Ann Rev Neurosci 24:167-202.

Nummela S, Jovanovic V, de la Mothe LA, Miller CT (2017) Social context-dependent activity in marmoset frontal cortex populations during natural conversations. . J Neurosci 37:7036-7047.

Perrodin C, Kayser C, Logothetis NK, Petkov C (2011) Voice cells in primate temporal lobe. Current Biology 21:1408-1415.

Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK (2008) A voice region in the monkey brain. Nature Neuroscience 11:367-374.

Plakke B, Diltz M, Romanski LM (2013a) Coding of vocalizations by single neurons in ventrolateral prefrontal cortex. Hearing Research 305:135-143.

Plakke B, Ng CW, Poremba A (2013b) Neural correlates of auditory recognition memory in primate lateral prefrontal cortex. Neuroscience 244:62-76.

Romanski LM, Averbeck BB (2009) The primate cortical auditory system and neural representation of conspecific vocalizations. Ann Rev Neurosci 32:315-346.

Romanski LM, Averbeck BB, Diltz M (2005) Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. J Neurophys 93:734-747.

Toarmino C, Wong L, Miller CT (2017a) Audience affects decision-making in a marmoset communication network. Biology Letters 13:20160934.

Toarmino CR, Jovanovic V, Miller CT (2017b) Decisions to communicate in the primate ecological and social landscapes. In: Psychological Mechanisms in Animal Communication (Bee MA, Miller CT, eds), pp 271-284: Springer Verlag.

Tsao DY, Livingstone MS (2008) Neural mechanisms for face perception. Ann Rev Neurosci 31:411-438.

Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. Science 311:670-674.

Yartsev MM (2017) The emperor's new wardrobe: Rebalancing diversity of animal models in neuroscience research. Science 358:466.

3 Mechanisms for communicating in a marmoset 'cocktail party'

3.1 Abstract

A key challenge of species that communicate with acoustic signals is parsing the voice of a single speaker amid a cacophony of conspecific vocalizations, known commonly as the Cocktail Party Problem (CPP). While the various perceptual and cognitive processes that can be employed to resolve the challenges of the CPP have been studied extensively in humans and some non-mammalian species, a notable paucity of experiments exist on the topic for nonhuman primates. Here we sought to bridge this gap by developing an innovative, multi-speaker paradigm comprised of five Virtual Monkeys (VM) whose respective vocal behavior was systematically manipulated to explicitly test how marmoset monkeys solve the CPP during natural communication. Results indicated that marmosets learned the identity of an interactive Target VM from amid a cacophony of vocalizations produced by the Distractor VMs, and that these monkeys employed a myriad of perceptual mechanisms including selective attention to effectively communicate in the various Cocktail Party environments. Furthermore, these results suggest that the acoustic structure of the species-typical long-distance contact calls itself is functionally significant for resolving the CPP suggesting a potential evolutionary relationship between signal design and audition in these primates. These results provide the first quantitative insight into dynamic mechanisms that support natural communication in a primate cocktail party.

3.2 Introduction

Our ability to effectively converse with others is often complicated by the co-occurrence of other speakers and other sources of acoustic interference, classically illustrated by the Cocktail Party Problem [CPP] (Cherry, 1953; Mcdermott, 2009). The seeming effortlessness with which audition solves the myriad of challenges inherent to the CPP belies the suite of sophisticated

mechanisms throughout multiple levels of the auditory system – including both peripheral and central processes - that must work in parallel for successful communication to occur in such environments (Pressnitzer et al., 2008). Because the challenges of communicating in noisy environments are nearly ubiquitous across a wide range of taxa, it is perhaps unsurprising that similar perceptual mechanisms to resolve these challenges are likewise evident (Bee and Micheyl, 2008). What is less well understood, however, is the role of the various neural substrates and circuits in the ascending auditory pathway to support these mechanisms. Human and nonhuman primates, for example, share the core architecture of the cortical auditory system that is distinct to our Order (Kaas and Hackett, 1998, 2000; Hackett, 2009; Kaas, 2010), but how these homologous substrates support the perceptual and cognitive mechanisms needed to communicate in cocktail parties is not clear because of a dearth of studies exploring these issues in our simian cousins, particularly at the behavioral level. To bridge this considerable gap, we developed an innovative, multi-speaker, interactive playback paradigm for marmoset monkeys that both simulates a natural cocktail party environment and offers experimental control to systematically manipulate characteristics of an acoustic and social landscape. This novel paradigm affords the powerful opportunity to explicate the various mechanisms and behavioral strategies employed to overcome these challenges in a species of nonhuman primate. Our aim here was not to determine psychoacoustic thresholds of the perceptual processes that support auditory scene analysis, but rather to explicate how these mechanisms and others are leveraged under real-world conditions to overcome the CPP for active communication in common marmosets (*Callithrix jacchus*).

Studies suggest that humans are able to resolve the challenges of communicating in multi-speaker environment using a handful of perceptual cues, including the spatial separation of

the speakers and the acoustic idiosyncrasies of individual voices (Darwin, 1997; Bronkhorst, 2015). Under natural conditions, listeners are typically given important cues that facilitate parsing an acoustic scene, including each persons' voice and the fact that each person is speaking from a distinct location in space. Even relatively small distances between speakers can increase intelligibility significantly while differences in each speaker's voice pitch provides a reliable cue (Bregman, 1994; Brungart and Simpson, 2002; Brungart and Simpson, 2007). In more dynamic scenes involving numerous speakers, these cues may become less clear, requiring listeners to employ more top-down perceptual mechanisms to selectively attend to particular speakers (Darwin and Hukin, 2000; Hill and Miller, 2009). During speech, one could learn a speaker's voice and segregate it into a single stream, potentially as a learned schema, facilitating its segregation from other sounds in the environment. While a handful of neurophysiological and behavioral studies in monkeys are suggestive that some auditory scene analysis mechanisms are used in primates (Miller et al., 2001; Micheyl et al., 2003; Petkov et al., 2003; Micheyl et al., 2005), there is a notable paucity of work explicating how nonhuman primates resolve the CPP. Certainly observations suggest that primates are able to communicate in noisy environments, but whether this is accomplished principally through bottom-up auditory mechanism or involves more top-down attentional processes similar to humans is not yet known (Shinn-Cunningham, 2008).

Common marmosets are a highly voluble New World monkey who naturally engage in conversational exchanges within natural communication networks reflective of the challenges of the Cocktail Party (Eliades and Miller, 2017). Like human conversations, the temporal dynamics of marmoset conversations are governed by learned social rules (Miller and Wang, 2006; Chow et al., 2015; Toarmino et al., 2017). The current study sought to address this

issue by building on our previous interactive playback paradigm (Miller et al., 2009b; Miller and Thomas, 2012) to construct Cocktail Party scenes using a multi-speaker design in which a single live monkey heard the vocalizations of five Virtual Marmosets (VMs) whose respective vocal behavior differed relative to the subject's. In this innovative design, the behavior of one VM – the Target – was designed to directly interact with the live marmoset, emitting vocalizations in response to the subject in order to engage them in conversational exchanges, while the timing of the other VMs – the Distractors - were independent of the subject. Calls from pairs of VM Distractors were structured to simulate a natural conversational exchange. This innovative paradigm afforded a powerful opportunity to systematically manipulate features of the acoustic scene (e.g. spatial separability and predictability of caller location, distractor density, and the acoustic structure of the vocalizations themselves) in order to explicitly test their effect on subjects' propensity to engage in conversational exchanges; thus providing key insights into the mechanisms that this nonhuman primate employs to overcome the challenges of communicating in a cocktail party environment. Importantly, the experiments here focus on how marmosets' propensity to engage in natural conversations were affected by these manipulations. This behavior represents an active communication exchange that requires the coordinated effort of two individuals to identify a willing partner amid a cacophony of conspecific vocalizations. In other words, it reflects a natural successful communication event during which the animals have successfully resolved the CPP and, therefore, provides a unique opportunity to illuminate mechanisms employed for that process.

3.3 Methods

3.3.1 Subjects

Six adult marmosets (3 females and 3 males) participated as subjects in these experiments from September 2019 to May 2020. All subjects were social housed in pair-bonded family units that comprised of two adults, and up to two generations of offspring. The UCSD Institutional Animal Care and Use Committee approved all experimental procedures.

3.3.2 Experimental Design

All experiments were performed in a ~4 X 3 m Radio-Frequency Shielded testing room (ETS-Lindgren). Individual subjects were transported from their home cage in clear acrylic transport boxes to the experimental chamber and tested individually. Subjects were placed in an acrylic and plastic mesh test cage (32 X 18 X 46 cm) designed to allow the animals to climb and jump freely along the front wall of the cage similarly to previous experiments (Miller and Thomas, 2012; Toarmino et al., 2017) . The cage was placed on a rectangular table against the shorter side of the room. Seven speakers (Polk Audio TSi100, frequency range 40-22,000 Hz) were placed on the opposite side of the room arranged to maximize distance relative to all other speakers in both the horizontal and vertical planes (Figure 1A). All vocal stimuli were broadcast at 80 dbSPL measured 0.5 m in front of the speaker. A cloth occluder divided the room to prevent the subjects from seeing any of the speakers during testing. One directional microphone (Sennheiser, model ME-66) was placed approximately 0.3 m in front of the subject to record all vocalizations produced during a test session. Another directional microphone was placed in front of the central speaker as well. We tested subjects three times to each test condition across two experiments while randomized. The order of each condition within the individual Experiments

was counterbalanced across subjects in a block design for the High and Low Distractor Density levels.

Cocktail Party Test Environments were constructed using an innovative multi-speaker paradigm in which vocalizations were broadcast from five, software generated Virtual Marmosets (VMs) (Figure 3.1A). The unique identify of each VM was determined by (1) broadcasting prerecorded vocalizations from an individual marmoset in the UCSD colony and (2) its vocal behavior relative to the live subject and other VMs. With respect to this later characteristic, VM vocal behavior was determined by their designation as a Target or Distractor. Similar to our previous experiments (Miller and Thomas, 2012; Toarmino et al., 2017), the behavior of Target VM was specifically designed to directly engage subjects in the species-typical natural conversational exchanges by utilizing an interactive playback design. To this end, the Target VM would broadcast a phee call response within 1-5s with an 85% probability each time subjects produced a phee call. In successive vocal exchanges between the subject and target (e.g. a conversational exchange), the Target VM would broadcast a response with 100% probability to maintain the vocal interaction. If subjects did not produce a call within 15-30s, the Target VM would broadcast a spontaneous call. Custom-designed software recorded vocal signals produced by the test subject from the directional microphone positioned in front of the animal and identified when subjects produced a phee call. By contrast, the timing of Distractor VM phee calls were independent of subjects' behavior, occurring at a predetermined interval. In each test condition, we generated two pairs of Distractor VMs. Each pair was designed to directly engaged each other in conversational exchanges. The timing of phee calls within these conversations was determined by the parameters of the test condition.

3.3.2.1 VM Stimulus Sets

All phee calls used as stimuli in these experiments were recorded from animals in the UCSD colony using standardized methods in the laboratory described in previous work (Miller and Thomas, 2012; Toarmino et al., 2017). Briefly, two monkeys were placed in separate testing boxes positioned ~3m from each other with an opaque cloth occluder located equidistant between the boxes to eliminate visual contact between the animals. Directional microphones (Sennheiser ME-66) were placed directly in front of each subject to record vocal output separately from each animal. Naturally produced calls were recorded direct to disk over a 30min sessions. At the conclusion of the session, custom-designed software was used to extract two-pulse phee calls produced during each session. Phee calls produced within 10s of a conspecific phee were classified as 'antiphonal' phee calls, while those produced after this threshold were classified as 'spontaneous' phee calls. These designations were based on previous research (Miller et al., 2009b). Each VM in a test session would only broadcast antiphonal and spontaneous phee calls from a single marmoset. The stimulus sets used as the basis for each Target and Distractor VM was randomized across test sessions. The VMs stimulus sets used to construct each Cocktail Party Scene were never from animals in a subject's home cage because of confounds that might occur due to social relatedness (Miller and Wang, 2006).

3.3.3 Test Conditions.

We selectively manipulated two dimensions of the acoustic and social landscape to directly test their respective impact on how marmosets resolved the challenges of communicating in a cocktail party in two experiments: *spatial configuration* & *distractor density*. Experiment 1 tested subjects using two-pulse phee calls as vocalization stimuli produced by VM, while Experiment 2 broadcast only 1-pulse phee calls from the VMs. To establish Baseline vocal

behavior in these Cocktail Party scenes, subjects were tested in the Fixed-Location using the same parameters as under normal conditions with one key exception. In these Baseline sessions, the Target VM calls were not broadcast to subjects. This allowed us to determine the probability that Target VM and subject would, through the natural statistics of marmoset vocal behavior in these scenes, occur in a temporal sequence consistent of conversational exchanges and compare it to subjects' behavior under conditions in which the Target VM were broadcast. A baseline condition was performed separately for High and Low Distractor Density levels and separately for both Experiments 1 and 2.

3.3.3.1 Spatial Configuration.

The spatial location of the VMs was manipulated by broadcasting the phee stimuli in three different speaker configurations: *Fixed-Location*, *Random-Location* and *Single-Location* (Figure 3.2A). These configurations allowed us to contrast the effects of both the significance of spatial separation between the callers and the predictability of a caller's position in space on marmoset vocal behavior.

3.3.3.1.1 Fixed-Location

In this configuration, the calls of each VM were broadcast from among five distinct, spatially separated speakers. This scene afforded subjects spatial separability of each VM from a consistent spatial location for the duration of the experiment.

3.3.3.1.2 Random-Location

Like the Fixed-Source condition, VM calls were broadcast from distinct spatially separated speakers. Rather than each VM broadcast from their own speaker for the duration of the experiment, speaker location was randomized across all 7 potential speakers during each broadcast. No VM call would be broadcast from the same speaker twice in a row, nor was there

any overlap in VM calls from the same speaker. As a result, subjects were afforded spatial segregation of the VMs, but with no predictability for where the VM would emit a call.

3.3.3.1.3 Single-Location

Here all VM stimuli were broadcast from a single speaker, thereby eliminating spatial separation of the different callers.

3.3.3.2 Distractor Density

Distractor density was manipulated to two levels – Low and High – by changing the relative inter-call interval between phees broadcast between VM Distractor pairs. In the 'Low' distractor density scene, Distractor VM conversations had an inter-VM call interval ranging 1 to 3.5 sec in Experiment 1 [2-pulse phee calls] and 1 to 2.5 sec in Experiment 2 [1-pulse phee calls]. In the 'High' distractor density scenes, Distractor VM conversations had an inter-VM call interval ranging from 0.5 to 1.0 sec in Experiment 1 [two-pulse phee calls] and 0.5-0.75 sec in Experiment 1 [one-pulse phee calls]. The shorter inter-VM call interval ranges for Experiment 2 were used to maintain the same level of Distractor Density when the shorter one-pulsed phee calls were used as stimuli.

3.3.4 Data Analysis

We calculated three behavioral metrics to quantify changes in subject vocal behavior relative to the Target and Distractor VMs as well as standard acoustic parameters, such as call duration and response latency.

3.3.4.1 Conversation Index

This metric quantified the ratio of subject's calls that were in conversation to all calls produced by the subject with said calls weighted by their position within conversations.

Previous experiments in marmosets determined that phee calls produced within 10s following a conspecific phee call were perceived as a 'response' to the initial call by conspecifics and were significantly more likely to elicit a subsequent vocal response, while those produced after this threshold did not elicit vocal responses from conspecifics (Miller et al., 2009b). Marmoset conversations are defined as instances in which monkeys engage in a series of alternating, reciprocal phee exchanges during which the inter-call interval between phee calls is within the 10s threshold. Each conversation ended when the subject did not respond for more than ten seconds. We elected to use conversations as our key behavioral metric in these experiments because they are indicative of learned communication behavior that requires a coordinated, interactive effort between marmosets (Chow et al., 2015).

For each test session, the temporal relationship between subjects calls and each of the five VMs was measured to determine whether the subject and a VM engaged in a conversational exchange. To quantify the occurrence of conversations in each test session, we first identified all instances in which the timing of subjects' phee calls and each VM conformed to these parameters using custom software. Subjects calls in conversational exchanges were assigned a number based on their linear order in the vocal exchanges sequence. In other words, the first response was assigned 1, the second successive response was assigned 2, etc. Standalone subject calls and the initiation of a conversational exchanges by subjects were assigned 0. By taking the average of all these calls across an experimental condition for each Subject-VM pair we can calculate their respective Conversation Index. This metric allowed us to compare the occurrence of conversations between the subject and each class of VM – Target and Distractors – as well as across test conditions.

3.3.4.2 Interference Ratio

We measured the temporal overlap between the Distractor VMs calls and the Target VM calls to determine the amount of acoustic interference that occurred. Each time a Target VM call was broadcast, we measured the duration of time it temporally co-occurred with any Distractor VM call. The resultant ratio indicates the percentage of overlap in time between Target and Distractor VM calls.

3.3.4.3 Pulse-Number Index

Custom software extracted all phee calls produced by subjects in each test session and identified the number of pulses within these calls based on previously identified stereotyped spectro-temporal structure of these vocalizations (Miller et al., 2010). Once cataloged, we could then compare the number phee calls produced that comprised 1, 2 or 3+ pulses. Previous studies have shown that the majority of marmoset phee calls consist of 2-pulses (~70%), while the other variants occur at lower frequency. Phee calls consisting of 3 or more pulse calls were rarely produced in the current experiments, accounting for <10% of calls, these were grouped together. Because the number of phee calls comprising 3+ pulses did not vary across the test conditions, these were excluded from this this metric. We generated the Pulse-Number Index by calculating the difference over the sum of the 1 and 2 pulsed phee calls produced in each session $[(1PulseRatio - 2PulseRatio)/(1PulseRatio + 2PulseRatio)]$. Positive values would indicate a bias towards 1-Pulse Phee calls, while a negative would reflect a bias towards 2-pulse Phee Calls.

3.3.5 Linear Model Analysis

3.3.5.1 Response Variables

These metrics were used as response variables within our linear models as mentioned in the Results section. Each one was calculated for each recorded session within a given experimental condition (18 per condition):

Average Duration of Calls: The mean duration of subject calls.
Duration of 1 Pulse Calls: Mean duration of 1-pulse calls produced by the subject
Pulse-Number Index: The difference over sum of the ratio of one pulse calls to two pulse calls produced by the subject.
Conversation Index: The mean position of the subject calls as previously mentioned.
Response Latency in Conversation: The mean latency of subjects to respond to Target VM within a conversational exchange.
Number of Calls: Number of calls produced by the subject in a given session.
Number of Conversations: The number of times the subject engaged in conversational exchanges.
Length of Conversations: The mean number of subject calls produced within each conversation.

3.3.5.2 Design

MATLAB function 'fitlm' was used to fit six predictor variables to each of the 8 response variables thus creating 8 linear models of comparison on 144 observations per model. The six predictor variables were: the calculated Interference Ratio (as seen in Figure 3.2B,C and Figure 3.3A,B), Distractor ICI, COV Distractor ICI, the categorical Distractor Density (Low or High), the categorical spatial configuration (Fixed or Single), and the categorical Experiment (2-Pulse or 1-Pulse). An interactive linear model was created that included an intercept term (1), linear term for each predictor (6), and products of pairs of distinct predictors excluding squared terms (15), for a total of 22 predictor terms. The 8 models created with 22 predictor terms were corrected for multiple comparisons using the Bonferroni correction. With a criterion at $\alpha = 0.05$, the new p-value threshold was calculated to be at $0.05/176 = 0.000284$. Any model's F-test for a degenerate constant model that was below this threshold was included for further analysis of the

terms. Four models reached this threshold as mentioned in the results. Of those four, only three had terms with coefficients that were significantly different from 0 below the corrected new threshold and were subsequently explored in Figure 3.4C-G.

3.4 Results

We tested 6 adult common marmoset monkeys in a series of experiments designed to examine the mechanisms that support communication in cocktail party environments. We completed two experiments comprising a total of fourteen different test conditions. Notably, we observed no statistically significant difference in the number of vocalizations produced by subjects across these condition (3-way ANOVA, total $df = 251$, $p = 0.693$). The mean amount of phee calls produced by each subject within each recording session per condition was 60.4 calls with a standard deviation of 34.3 calls. This suggests that none of the cocktail party landscapes constructed in these experiments suppressed marmoset vocal behavior. Rather, subjects consistently attempted to engage with VMs in communicated exchanges throughout.

In these experiments, we quantified a series of behavioral metrics to determine how challenges of the cocktail party affected marmoset vocal behavior. First, we determined whether marmosets could communicate in different social landscapes by calculating a 'Conversation Index'. This metric quantified the propensity of subjects to engage in their naturally occurring conversational exchanges, characterized by the reciprocal exchange of phee calls (Miller and Wang, 2006; Miller et al., 2016). Briefly, each of subjects' vocal response to VM calls were given increasing incremental values that corresponded to their successive position within the conversation. This was averaged to generate a Conversation Index for each condition. See Methods for an expanded description. Second, we characterized how different dimensions of

marmoset vocal behavior changed as a function of specific manipulations of the cocktail environment. These detailed analyses are described below.

3.4.1 Experiment 1

Experiment 1 was designed to test how manipulation of the cocktail party environment along two axes affected marmoset conversations: Distractor Density and Spatial Configuration of the Location (Figure 3.2A). Each of these features of the acoustic scene are known to influence how humans resolve the CPP (Bronkhorst, 2015). We hypothesized that, like humans, these perceptual challenges would likewise affect marmosets' capacity to communicate in cocktail party environments.

The results from Experiment 1 are shown in Figure 3.2. These experiments broadcast 2-Pulse phee call stimuli from each VM at two Distractor Density levels – High and Low – in three spatial configurations – Fixed-Location, Random-Location, and Single-Location (Figure 3.2A) - as well as the Baseline condition. Importantly, the Baseline condition differed from the other test conditions in a critical way. Here, the Target VM vocalizations were not broadcast. Rather, the system would record the timing of the stimulus, but the vocalization would not be broadcast from a speaker. This condition served to identify the baseline volubility and call timing of subjects' in the absence of any interactive feedback from the Target VM, but in the same cocktail environment. The condition was crucial because the natural spontaneous call rates of marmosets could result in response false positives (Miller and Wang, 2006). We compared subjects' Conversation Index in the Baseline Condition across the other Spatial Configuration conditions to determine whether the propensity of marmosets to engage in conversations statistically differed when interactive feedback from the VM occurred. Baseline conditions were performed separately for each Distractor Density level.

Figures 3.2B and 3.2C, plot the 'Low' and 'High' Distractor Density levels, respectively. Distractor Density was calculated as the ratio of the Target VM calls that temporally overlapped with Distractor VM calls. Figure 3.2B shows that the 'Low' Distractor Density had a mean interference ratio of 0.724 across all the sessions. In other words, on average,72% of the duration of the Target VM calls broadcast acoustically overlapped with one or more of the Distractor VMs calls. By contrast, the mean interference ratio for 'High' Distractor Density in this experiment was 0.903 (i.e. 90%, Figure 3.2C). The standard error was 0.00358 for both Distractor Density levels. At both Distractor Density levels, the Conversation Index of subject calls with respect to Target VM was significantly higher than the Distractor VMs (2-way ANOVA, total df = 43001, VM df = 4, $p = 0$). The Distractor VMs had means at 0 while Target VM call index averaged at 1.30. As a result, we did not further explicate Distractor VM conversation index values for analysis.

Figure 3.2D shows subjects' Conversation Index for Baseline and the three spatial configurations of Fixed-, Random-, and Single-Location at the Low Distractor Density level. The mean of the Baseline was subtracted from the three spatial configurations to represent the relative change from Baseline. Although subjects exhibited a significant broad increase in Conversation index across all conditions relative to Baseline (2-Way ANOVA (Spatial and VM), total df = 22934, Spatial*VM df = 12, $p < 0.0001$), this was largely driven by two of the three spatial conditions. For both the Fixed and Single location conditions, subjects' Conversation Index was significantly higher relative to baseline (Tukey-Kramer multiple comparison corrected $p = 0.0001$ and $p < 0.0000$, respectively), but not the Random-Location condition. The Conversation Index for Fixed and Single were above Baseline by at 0.200 and 0.267. These data suggest that at this Distractor Density, marmosets were able to correctly identify the Target VM

and selectively engage them in conversational exchanges. The pattern of subjects' behavior across these conditions suggests that spatial configuration of the VMs does play a role in resolving the CPP in these settings, but not necessarily the separability of the VMs in space. Rather, it is the predictability of the Target VMs location that is crucial as it affords an advantage.

Figure 3.2E plots subjects' Conversation Index for the same conditions as above, but at the High Distractor Density level. Notably, marmosets' behavior was similar despite increased interference from the Distractor VM. As in the low Distractor Density scenes, marmosets exhibited overall higher calling to the Target VM relative to Baseline (2-way ANOVA (Spatial and VM), total df = 20094, Spatial*VM df = 12, $p < 0.0001$). However, comparison of individual conditions revealed that subjects' Conversation Index was significantly higher than Baseline only for the Fixed and Single conditions (Tukey-Kramer corrected $p<0.0001$ for each). The respective Conversation Index for Fixed and Single were above Baseline by 0.410, 0.430, respectively. In contrast to the Low Distractor Density, however, Conversation Index for the Random-Location condition was significantly lower than the Single or Fixed conditions (Tukey-Kramer corrected $p < 0.0000$ for both). The broad similarity between the High and Low Distractor densities in this experiment suggest that the increase in interference by conspecifics did not significantly impair marmosets' ability to resolve the challenges of communicating in a cocktail party, as long as the location of the Target VM was in a predictable location. This suggests that focusing attention to a spatial position may help to offset the challenge of parsing the Target VM from the Distractors when acoustic interference was nearly omnipresent. Given the notably slow periodicity of marmoset conversational exchanges, with inter-call intervals up

to 10s in duration, leveraging this mechanism may be crucial to resolving the challenges of communicating in a Cocktail Party.

The ability of marmosets to correctly identify the Target VM and selectively engage with them in conversational exchanges irrespective of Distractor Density was somewhat surprising. However, marmosets effectively engaging in conversational exchanges does not reveal the more detailed nuances of how they accomplished this feat. We next performed a series of analyses to determine whether more nuanced facets of their vocal behavior differed between the test conditions. As shown in Figure 3.2F, subjects produced a lower ratio of 1 pulse calls at the High Distractor Density level 55.2% to 47.6%, while 2 and 3 pulse calls modestly increased (2 pulse +6.70%, 3+ pulse +0.980%); a pattern found to be statistically significant (Kruskal-Wallis test, $df = 6453$, $p < 0.0001$). Notably, the median phee call variant produced in these experiments changed from 1-pulse phee at the Low Distractor Density to 2-pulse phee calls at the High Distractor Density. Figure 3.2G further shows that there was a significant change in both the average duration of phee calls (+9.12%), and the 1 pulse phee calls (+9.98%), but not 2 or 3+ pulse phee calls, from Low to High Distractor Density (3-Way ANOVA (Spatial, Acoustic, Pulse count), total $df = 6453$, acoustic $p < 0.0001$ and acoustic*pulse count $p < 0.0001$). Finally, we next compared the latency that subjects responded to the Target VM within conversations at the two Distractor Density levels (Figure 3.2H). Analyses indicated that the distribution of this latency was significantly shorter at the higher acoustic interference level (307ms; Kruskal-Wallis test, $df = 2714$, $p = 0.0207$). Together these results indicate that subjects increased the median duration of their phee calls and decreased the latency to respond to Target VMs when communicating at higher Distractor Density. This change in vocal behavior strategy may have

been necessary to maintain conversational exchanges in Cocktail Parties with a near constant levels of acoustic interference from conspecifics.

3.4.2 Experiment 2

The acoustic structure of long-distance contact calls – including the marmoset phee call (Morrill et al., 2013) – has been selected over evolution to maximize signaling efficacy in noisy environments (Waser and Brown, 1986; Mitani and Stuht, 1998). To this end, a common characteristic of this class of vocalizations is the repetition of an acoustically similar pulse. Such acoustic redundancy is speculated to have evolved because it functions to offset the inherent decline in the acoustic content of the vocalizations as they travel over long distances (Waser and Waser, 1977). Marmoset phee calls are consistent with this trend, comprising a series of acoustically similar repeated pulses (Miller et al., 2010). In Experiment 1, all VMs emitted 2-pulse phee calls because this is the most common variant of the call, accounting for nearly 70% of phee calls produced by marmosets (Miller et al., 2010). Here we tested whether this signal design characteristic was beneficial to marmosets communicating in a Cocktail Party. We hypothesized that if redundancy in call structure was beneficial to marmosets, eliminating this characteristic of the call would result in increased difficulty maintaining conversational exchanges. We tested subjects in the same Environments as in Experiment 1 but used 1-pulse phee calls as the stimulus produced by each VM rather than the 2-pulse phee calls used in the previous experiment. Given that subjects already struggled to communicate in the Random-Location condition under less challenging conditions, we did not repeat this test condition here. Instead we tested subjects only in the Fixed-Location and Single-Location spatial configurations.

Results for the Low Distractor Density in Experiment 2 are shown in Figure 3.3C. Consistent with Experiment 1, the Conversation Index for the Distractor VMs was significantly

lower than the Target VMs in Baseline, Fixed, and Single conditions (Tukey-Kramer corrected, $p < 0.0000$), while all the distractors amongst each other had no difference (Tukey-Kramer corrected, $p = 1.0000$). A comparison of Conversation Index with the Target VM between the test conditions showed that subjects engaged in significantly more conversations in both the Single and Fixed conditions relative to Baseline (2-Way ANOVA Spatial x VM, total df = 16579, $p < 0.0001$), with Fixed and Single source conditions having a mean Conversation Index above Baseline by 0.672 and 0.472 (Tukey-Kramer corrected, $p < 0.0001$ for both), respectively. In contrast to parallel results in Experiment 1, the Fixed and Single source conditions were statistically different from each other (Tukey-Kramer corrected, $p = 0.0379$). These results suggest that, although marmosets could identify the Target VM and maintain conversational exchanges in both conditions, the spatial separation between the various VMs in the Fixed-Location condition may have afforded some perceptual advantages despite the spatial predictability when only hearing 1-pulse phees emitted by the VMs even at the Low Distractor Density level.

A comparison of Conversation Index across the test conditions at the High Distractor Density is shown in Figure 3.3D. Similar to previous conditions, there was no difference in Conversation Index between the Distractor VMs (Tukey-Kramer correction $p = 1.00$) but a significant difference in the Target VM conversation indexes for Baseline, Fixed-Location, and Single-Location (Tukey-Kramer correction $p < 0.0001$). We observed a significant interactive effect of spatial configuration (2-Way ANOVA Spatial x VM, total df = 16434, $p < 0.0001$), but here only the Fixed-Location was statistically significant from Baseline. The Fixed condition had significantly higher mean Conversation Index at 0.490 above Baseline (Tukey-Kramer corrected, $p < 0.0001$), while the Single source condition was below Baseline by -0.0788 (Tukey-Kramer

corrected, $p = 0.994$). These results suggest that when the Cocktail Party consists of near constant acoustic interference from conspecifics producing only 1-pulse phee calls, the advantages afforded by selectively attending to a predictable location in space alone was insufficient for marmosets to consistently identify and engage in conversational exchanges with the Target VM. Rather, spatial separability of the callers was critical to maintain effective communication when the acoustic content of the vocal signal was limited to a single pulse.

We next analyzed how the challenges of communicating in these Cocktail Party Environments affected subjects' vocal behavior. Analyses revealed that like Experiment 1, marmosets systematically modified their communication behaviors. The pattern of changes, however, were notably different from what we observed in the previous experiment. Figure 3.3E shows that there was a significant change in the distribution of the number of pulses per call made by the subject (Kruskall-Wallis, $df = 4424$, $p < 0.0001$). Whereas here we observed a higher ratio of 1 pulse calls produced by the subjects in the High Distractor Density conditions, Experiment 1 had the opposite effect. The ratio of 1 pulse calls produced by subjects increased from 59.6% to 70.7%, the 2 pulse and 3+ pulse calls dropped (-11.1% and -0.06%). Again, in contrast to results seen in Experiment 1, Figure 3.3F shows this did not result in a significant overall change in the duration of calls produced by subjects; rather, the significant changes in duration was apparent when subject calls were broken down by the number of pulses (3-Way ANOVA (Spatial, Acoustic, Pulse count), total $df = 4424$, acoustic $p = 0.385$ and acoustic*pulse count $p < 0.0001$). The 1 pulse calls increased in duration from lower to higher by 10.6% (Tukey-Kramer corrected $p < 0.0001$). The 2 pulse and 3+ pulse calls did not change significantly from lower to higher at -1.17% and 7.04%, respectively. (Tukey-Kramer corrected $p = 0.427$ and $p = 0.788$). We also observed a significant decrease in latency to respond to Target

VM at the High Distractor Density level relative to the lower level (Kruskall-Wallis test, df = 2119, p = 0.0019), similarly to Experiment 1 (Figure 3.3G). The median latency within the conversation went from 5.09 sec to 4.71 sec in lower to higher conditions (-7.54% change or 384 msec shorter response time). One potential reason for the notable change to producing more 1-pulse phee calls in this experiment may be because of an unintended change in the acoustic scene statistics that emerged as a result of matching the amount of acoustic interference between the two experiments. In doing so, the inter-call interval between the VM Distractor calls decreased significantly which may have driven marmosets to adjust their own call structure to compensate for the increased periodicity of the conversational exchanges, as has been observed previously in marmosets (Roy et al., 2011). This observation suggests that marmosets may have implemented more adaptive changes to vocal behavior in response to the dynamics of the acoustic scene that emerged as a biproduct of the Cocktail Party landscapes generated in these experiments.

3.4.3 Emergent Acoustic Scene Dynamics Reveal Adaptive Changes in Vocal Behavior

An unintended biproduct of our effort to control for acoustic interference across the two experiments was systematic changes to other dimensions of the acoustic scene statistics. Most notably was a systematic change in the timing and variability of the interval between the Distractor VM calls to achieve the desired Distractor Density when constructing these conversations. To explore their respective impact on marmoset vocal behavior, we limit our analyses only on the Fixed-Location and Single-Location conditions because the Random-Location was not performed in Experiment 2.

Figure 3.4A shows the distribution of the mean inter-call interval (ICI) against the calculated Distractor Density for each session within Fixed-Location and Single-Location (Low and High Distractor Density for both Experiment 1 and 2). Notably, significant negative

correlations exist between the two values for both Experiment 1 and 2 (rho = -0.796 & $p < 0.0000$, rho = -0.935 & $p < 0.0000$, respectively). The broad pattern revealed by these quantifications emerged because the shorter duration 1-pulse phee calls necessitated a shorter ICI between VM distractor pairs to ensure similar levels of Distractor Density across the experiments. This characterization formed the foundation for the subsequent statistical analyses aimed at explicating the relationship between the emergent scene structure and marmoset vocal behavior in these experiments.

We next applied a linear model to test how facets of marmoset vocal behavior covaried with dimensions of the acoustic scene. The following were input into the Linear Model - VM Pulse # (2-pulse:Expt 1, 1-pulse:Expt2), Low and High Distractor Density, and Fixed and Single conditions – for a total of 144 sessions. We also chose to include the calculated Distractor Density for each session along with the Distractor ICI. Given a strong positive correlation between Distractor ICI and standard deviation (rho = 0.931 and $p < 0.0001$), we took the coefficient of variance (COV, standard deviation divided by mean) as a way to encapsulate these two correlated factors while avoiding rank deficiency in any linear model (COV v Mean ICI, rho = -0.0956, $p = 0.254$. Figure 3.4B). This also gave an added benefit of enumerating the relative dispersion of the Distractor ICI. This analysis yielded six total predictor variables. The following 8 vocal behavior response variables were also input into the GLM: the mean duration of all calls, the duration of the 1-Pulse calls, Index of relative 1 and 2 pulse calls produced by subjects (Pulse Number Index), the Conversation Index, subjects mean latency to respond in a conversation, the number of subject calls produced, the number of conversations, and the mean length of those conversations.

We tested eight interactive linear models which included 22 terms (1 intercept, 6 linear predictor terms, and 15 pairs of distinct predictor terms). The statistical threshold for significant terms and models was corrected for multiple comparisons with Bonferroni correction based on 22*8=176 comparisons with a corrected P value threshold at 0.05/176 = 0.000284. Of these eight models, four models reached significance: duration of the subject 1-Pulse calls, Pulse Number Index, number of subject calls, and number of conversations (R^2 = 0.332, 0.396, 0.451, 0.489, adjusted R^2 = 0.216, 0.291, 0.357, 0.401). Of these four models, the duration of 1-Pulse subject calls did not have a significant term below the corrected threshold. Two significant terms were shared across the remaining three significant models. The Distractor ICI x COV Distractor ICI (which results in standard deviation Distractor ICI) for subject calls produced and number of conversations, and the mean distractor ICI x 1/2 Pulse VM Calls Condition for all three models. Figure 3.4C-G plots the five significant terms against the respective response variables in interaction effects plots. Each image plots the adjusted response function of the given response variables on the Y-axis against the values of the first predictor in the interactive term with the second predictor at fixed values (for categorical: all levels, and numeric: minimum, maximum, and average of minimum and maximum). Given that all five interactive terms have significant coefficients within their respective models, and that the slopes of the lines in all five plots are not parallel, there is significant interactive effect between the predictors for predicting the Pulse Number Index, the number of subject calls produced, and number of conversations.

Presenting subjects with VM calls comprising either 2 or 1 pulse phee calls – Experiments 1 and 2, respectively – resulted in opposite effects on the adjusted response variables. For Pulse Number Index (Figure 3.4C), Calls Produced (Figure 3.4E) and Conversation Count (Figure 3.4E), these behavioral metrics revealed a positive correlation with

Distractor ICI in Experiment 1, but a negative relationship in Experiment 2. In other words, when hearing 2-pulse VM calls in Experiment 1, subjects were more likely to produce 2-pulse phees, produce more calls and engage in more conversations as the Distractor ICI increased in duration. By contrast, the opposite was true when hearing only 1-pulse phee calls in Experiment 2. This suggests that the strategy to optimize communication exchanges when the Distractor streams are heard at a certain interval is not static, but highly correlated with the types of calls marmoset subjects heard in the Cocktail Party. In other words, the behavioral strategy did not change linearly as a function of the VM call rate, but that call rate was perceived to warrant different vocal behaviors from marmosets depending on which phee variant they heard in the scene. It should be noted that the consistency in the change of call rate and number of conversations is notable because these need not necessarily be parallel. It suggests that the increased number of calls are specifically being committed to conversations rather than calls produced independent of these active communicative exchanges.

A further significant factor affecting marmoset vocal behavior in the linear model was COV Distractor ICI. Both the number calls produced (Figure 3.4F) and conversations (Figure 3.4G) showed a similar pattern relative to Distractor ICI. As the Distractor ICI increased, at low COV, the relationship of number of subject calls and conversations produced decreased. At the highest level of COV, the opposite relationship emerged with increasing calls produced and conversations (with a smaller relative change). This suggests that as the predictability of the Distractor ICI increased (high to low COV), shorter Distractor ICI were optimal for the subject to produce calls and engage in more conversations with the Target VM. Similarly to the importance of spatial predictability for marmosets in Experiment 1, temporal predictability was

advantageous for marmosets to navigate the complex acoustic scene and selectively engage with the Target VM.

3.5 Discussion

Here we leveraged the advantages of our innovative, multi-speaker virtual monkey (VM) paradigm to systematically manipulate specific features of the acoustic landscape to test which mechanisms support nonhuman primate communication in Cocktail Party environments. Results clearly demonstrate that marmoset monkeys were readily able to identify a conversational partner and maintain communicative exchanges despite the complex acoustic and social landscapes comprising multiple conspecifics. A crucial question, however, pertains to how precisely marmosets solved the challenges of these test environments to maintain conversational exchanges. One possibility is that marmosets are simply excellent acoustic scene analyzers that are readily able to parse meaningful signals from interfering background masking noise principally using bottom-up mechanisms (Aubin and Jouventin, 1998; Bee and Micheyl, 2008; Bee, 2015; Lee et al., 2017). Certainly, such scene analysis mechanisms supported marmosets here, but those mechanisms alone cannot account for the pattern of results that emerged from these experiments; rather, evidence suggests that selective attention was likely used under at least some conditions to segregate the Target VM stream from the Distractors and resolve the CPP. Furthermore, a systematic failure to maintain conversational efficacy when hearing 1-pulse – rather than 2-pulse – phee calls is suggestive that the signal structure of this long-distance contact call may have evolved to facilitate this cognitive process for effective communication.

Evidence from the experiments here support the use of auditory attentional mechanisms by marmosets to resolve the CPP. Experiments in humans that employed a task involving multiple speakers found that when the spatial position of each talker randomly changed across

locations, subjects' intelligibility scores decreased (Brungart and Simpson, 2007). Likewise, human subjects performed significantly better when the spatial location of the target was cued prior to hearing the sound (Kidd et al., 2005). In both cases, it was concluded that the predictability of a talker's position in space allowed subjects to focus attention to that position in space. When that predictability was eliminated, attention could not be focused, and it accordingly had a negative impact on humans capacity to understand what was spoken. Experiment 1 sought to test whether a nonhuman primate would exhibit a similar pattern of behavior under these conditions. In the Random-Location condition, we randomized the spatial location across seven speaker locations each time one of the Target/Distractor VM calls were broadcast. Importantly, the vocal behavior of the VMs (i.e. the acoustic scene) was identical across all three test conditions, and the only difference in the Random-Location condition was the randomness of where each phee call was broadcast from amongst the seven speakers. As a result, marmosets could not predict where in the scene the Target VM call would be broadcast. As shown in Figure 3.2D&E, subjects performed significantly worse under these test conditions than either of the other two conditions. In fact, their conversational behavior was statistically indistinguishable from Baseline at both Distractor Density levels suggesting that a lower level of acoustic interference did not allow them to overcome the lack of predictability in the Target VMs spatial location. By contrast, marmosets were readily able to engage in conversational exchanges when VM calls were broadcast from a separate, but consistent locations like in Fixed-Location or where all VM calls broadcast from a single speaker like in Single-Location. The similarities between marmosets and humans (Kidd et al., 2005; Brungart and Simpson, 2007) when a talker's spatial position could and could not be predicted was nearly identical and suggestive that similar selective attentional mechanisms were leveraged to resolve the CPP in both primate species.

The dynamics of marmoset conversations may lend itself to a schema-based mechanism for speaker stream segregation (Bregman, 1994; Bey and McAdams, 2002; Woods and McDermott, 2018). First, marmosets needed to learn the identity of the Target VM for each session. While the spectro-temporal structure of marmoset phee calls is relatively stereotyped, lending itself to potential specializations for parsing the signal from the myriad of potential acoustic interference in the natural environment (Mcdermott, 2009), each monkey's phee is individually distinctive (Miller et al., 2010; Miller and Thomas, 2012). As a result, segregating one caller's phee call from amongst the phee calls of many conspecifics presents a different challenge that relies on learning the identity of a willing conversational partner. Although all calls broadcast from VMs were produced by animals in the UCSD colony who were familiar to subjects, no phee calls from subjects' cage-mates were used. As a result, familiarity was likely consistent across the VM callers (Johnsrude et al., 2013). Second, this learning occurred only based on direct feedback of subjects own vocal behavior. While subjects heard high number of calls from Distractor VMs, the lower call count in Target VM occurred interactively with subjects. The timing of Target VM calls conformed to the statistics of natural marmoset conversational exchanges and was designed to broadcast in response to subject's call as an interactive vocal exchange (Miller and Wang, 2006; Miller et al., 2009b; Toarmino et al., 2017). Therefore, marmosets learned the identity of the Target based on the statistical occurrence of this critical temporal cue in the VM's behavior relative to their own rather than anything intrinsic to the vocalizations themselves. Third, once the Target VM identity was learned, marmosets continuously monitor conspecifics' behavior and must temporally coordinate their own behavior for conversations to occur. A series of previous studies has shown that these temporal cues are governed by social rules that differ based on the age, sex and relatedness of the individuals and

are crucial for the coordination at the core of these vocal interactions (Miller and Wang, 2006; Miller et al., 2009b; Chow et al., 2015; Toarmino et al., 2017). Amongst these temporal dynamics, however, is the relatively slow periodicity of these conversations. Marmosets abide turn-taking in these conversations but the interval between calls is ~3s, but can range up to 10s (Miller and Wang, 2006; Miller et al., 2009b). Because of the cacophony of marmoset phee calls broadcast in these experiments, particularly at the high Distractor Density level, focusing attention to a predictable spatial location would have been notably advantageous considering a latency of several seconds between the offset of the subjects call and the Target VM response, during which time the calls of multiple Distractor VMs could have been emitted. Notably, however, the advantages of attention and a schema-based learning to stream the Target VM did have its limits, as evidenced by the difficulties of marmosets to communicate in the Single-Source condition for Experiment 2 when hearing only 1-pulse phee calls produced by VMs.

Results from Experiment 2 contrasted with Experiment 1 in several important ways that may reveal an evolutionary relationship between vocal signal design and audition in marmosets. While marmosets performed similarly in the Fixed or Single-Location test conditions when hearing 2-pulse phee calls from all VMs in the first Experiment, only showing difference at the higher Distractor Density level (Figure 3.2C&E). this pattern did not replicate in Experiment 2. When hearing only 1-pulse phee calls from VMs, marmosets did indeed continue to engage in conversational exchanges with the Target VM in the Fixed-Location condition irrespective of the level of acoustic interference, but exhibited notable declines when all VM calls were broadcast from a single speaker in the Single-Location condition (Figure 3.3B&D). Indeed, though marmosets engaged in conversations at significantly higher levels than Baseline in the Single-Location condition at the lower Distractor Density level, their performance statistically declined

relative to the Fixed-Location. Moreover, at the higher Distractor Density level, marmosets' conversations were statistically indistinguishable from Baseline. In other words, under these conditions spatial-release from masking was necessary to identify the Target VM and maintain conversational exchanges (Litovsky, 2005; Jones and Litovsky, 2011; Pastore and Yost, 2017). An explanation for this pattern likely pertains to the selective pressures on the phee calls themselves that occurred over evolution to maximize signaling efficacy (Morrill et al., 2013). Nonhuman primate long-distance contact calls – including the phee – often comprise the repetition of a single syllable, a signal design structure conjectured to limit degradation of the signals communicative content when transmitting long distances through noise acoustic environments (Waser and Waser, 1977). By effectively reducing the number of pulses in each call, we effectively halved the amount of acoustic information available to both identify the Target VM and recognize it in subsequent potential interactions. Indeed, in tamarin monkeys – a close phylogenetic cousin with marmosets – reducing the number of pulses in their contact call significantly impairs their ability to recognize the caller's identity (Miller et al., 2005). While marmosets were able to overcome this challenge when each VM called from a different - but consistent - position in space, subjects had difficulty (i.e. Low Distractor Density) or were unable to converse with the Target VM (i.e. High Distractor Density) when all calls were broadcast from a Single-Location. This suggests that under conditions with the highest acoustic interference, the redundancy of a two-pulse phee call is crucial to maintaining active conversational changes. Selection for multi-pulsed phee calls in marmoset evolution, and more broadly for other nonhuman primates, may have been driven specifically by the limits of audition for parsing vocalizations and recognizing callers amid the myriad of biotic and abiotic noise common in the species forest habitat.

We have thus far only considered the role of audition in resolving the challenges of communicating in Cocktail Party environments, but evidence here demonstrates that marmosets actively modified their own vocal behavior in response to changes in the acoustic scene statistics that emerged as a biproduct of how the environments were constructed. To control for acoustic interference, it was necessary to decrease the inter-call interval (ICI) between phees in the Distractor VM conversations. The emergent effect on the acoustic scene was a systematic change in properties of the conversation's periodicity – i.e. variance and inter-call interval. The effect of these scene characteristics on marmoset vocal behavior was considerable. When Distractor VM Conversations comprised 1-pulse calls and occurred at a faster call rate (i.e. decrease in Distractor ICI), marmoset had a propensity to produce more calls in more conversations while producing more 1 pulse calls (Figure 3.4C,D). By contrast, when Distractor VM conversations comprised 2-pulse phee calls, these same measures of volubility increased for slower Distractor VM conversations (i.e. increase in Distractor ICI). Furthermore, marmoset behavior was significantly influenced by the predictability of the Distractor ICI, as subjects exhibited significant biases to produce more calls (Figure 3.4F) and conversations (Figure 3.4G) when Distractor VM conversations were the least variable. These patterns are notable for several reasons. First, across the Linear Model, changes in the number of calls produced and conversations occur in parallel with Distractor VM ICI. This suggests that marmoset are not simply calling more, but are specifically producing more calls within the context of active communication. Second, interactive effects revealed by the model suggests that optimizing the dynamics of these conversations necessitates a dynamic strategy, one in which marmosets must exert control over both the call structure and the behavior (Pomberger et al.; Miller et al., 2009a; Pomberger et al., 2019; Zhao et al., 2019). Consistent with previous experiments (Roy et al.,

2011) and results in Experiment 1, the predictability of the environment plays a significant role in how marmosets resolve the CPP for effective communication. Finally, the outcome of this model suggests that subjects are not ignoring the Distractors VMs. Rather marmosets appear to attend to the dynamics of the Distractor VM conversations and actively modify their own behavioral strategies that optimize conversations with the Target VMs.

Here we show the mechanisms employed by a nonhuman primate – common marmoset monkeys – to actively communicate in Cocktail Party environments. These novel insights were possible because the innovative, multi-speaker paradigm developed for these experiments afforded the powerful opportunity to systematically manipulate various features of the acoustic scene in a manner not previously possible. As an inherently interactive process, communication is particularly at risk of decreased signaling efficacy in dynamic acoustic environments. By explicitly testing marmoset conversational exchanges, a coordinated vocal interaction, our experiments revealed significant insight into the mechanisms used to resolve the CPP in a primate. Results are suggestive that human and nonhuman primates likely resolve the CPP using similar mechanisms and lay a critical foundation for further explication of these issues at the neurobiological level. The neural basis of the auditory scene analysis and the CPP are poorly understood in primates. The marmoset auditory system shares the core functional architecture of all primates, including humans (Kaas and Hackett, 2000; de la Mothe et al., 2006; Bendor and Wang, 2008; Hackett, 2009), and has been a key primate model of sound processing, including vocalizations, for many years (Wang and Walker, 2012; Wang, 2013; Miller et al., 2016; Song et al., 2016 ; Eliades and Miller, 2017). By integrating existing technologies for recording neural activity in freely-moving marmosets with the current behavioral paradigm, the potential to

explicate the neural basis of the CPP in the primate auditory system with cellular resolution can be realized.

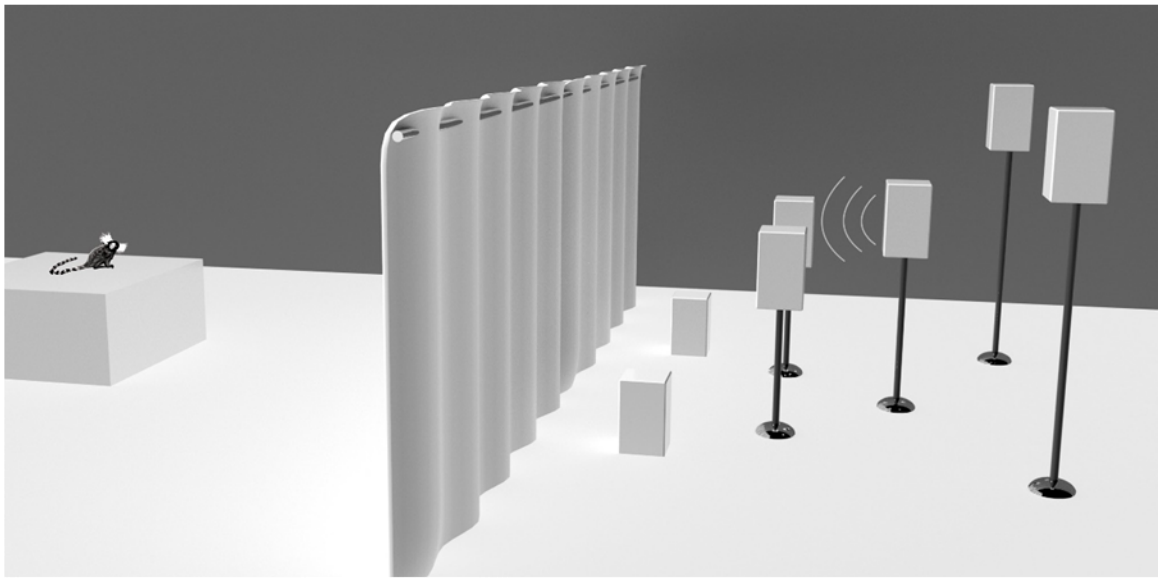
3.6 Acknowledgements.

This work supported by grants from NIH (R01 DC012087) and DARPA (SSC-5029) to C.T.M.

3.7 Figures

Figure 3.1: Design of the marmoset Cocktail Party experiments. (A) Schematic drawing of the spatial configuration of the testing room. Subjects were placed in a clear acrylic box with a mesh front (box around subject not pictured). Seven speakers were positioned to have spatial separation in height, distance and width. An opague curtain was placed equidistant between the subject and speakers to occlude visual access. (B) An exemplar two-minute sample of the vocalizations broadcast by the Virtual Monkeys (VM) and a live marmoset subject from a High Distractor Density, Fixed-Source session in Experiment 1. VM 1-4 are Distractors. VM1 and VM2 (shown in red) have been designed to broadcast 2-pulse phee calls that reflect a conversation with each other, while VM3 and VM4 (shown in brown) are likewise designed to engage in a reciprocal conversational exchange. The Target VM (blue) is engaged with the live marmoset Subject in an interactive reciprocal exchange based on subjects' vocal behavior. engaging in back The combined view shows the summation of all VM phee calls – Distractors (purple) and Target (blue).

A



B

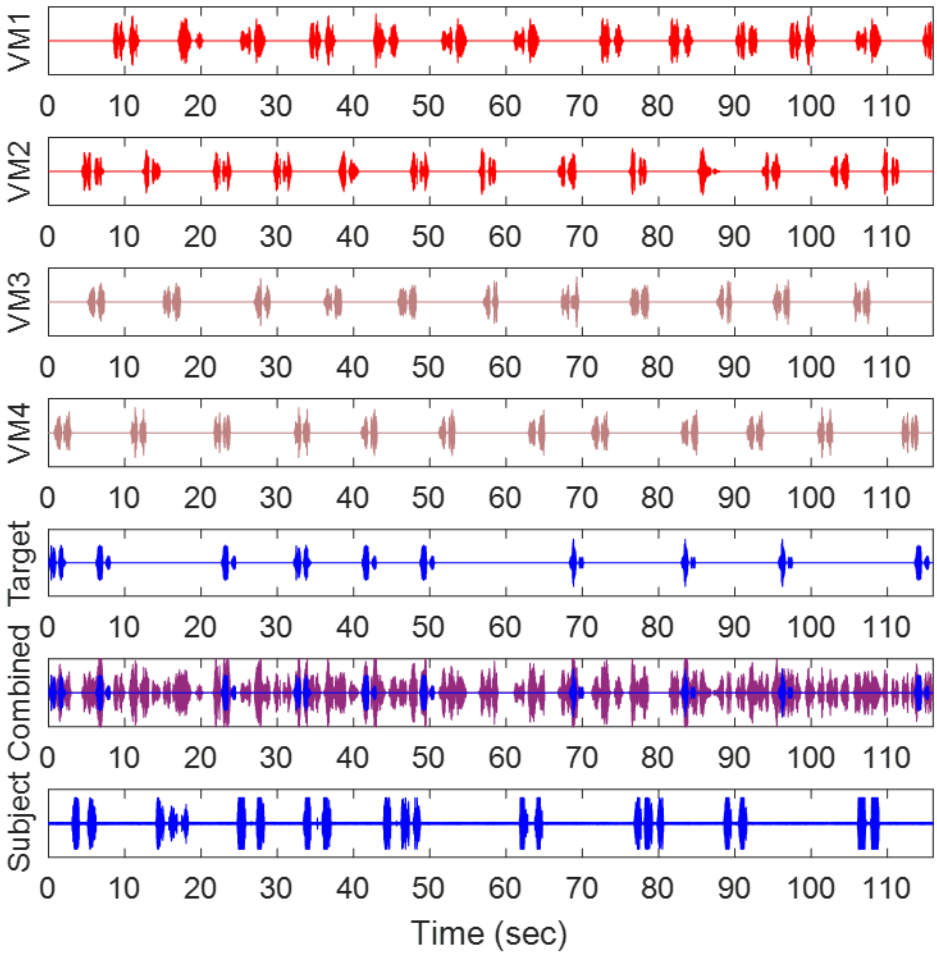
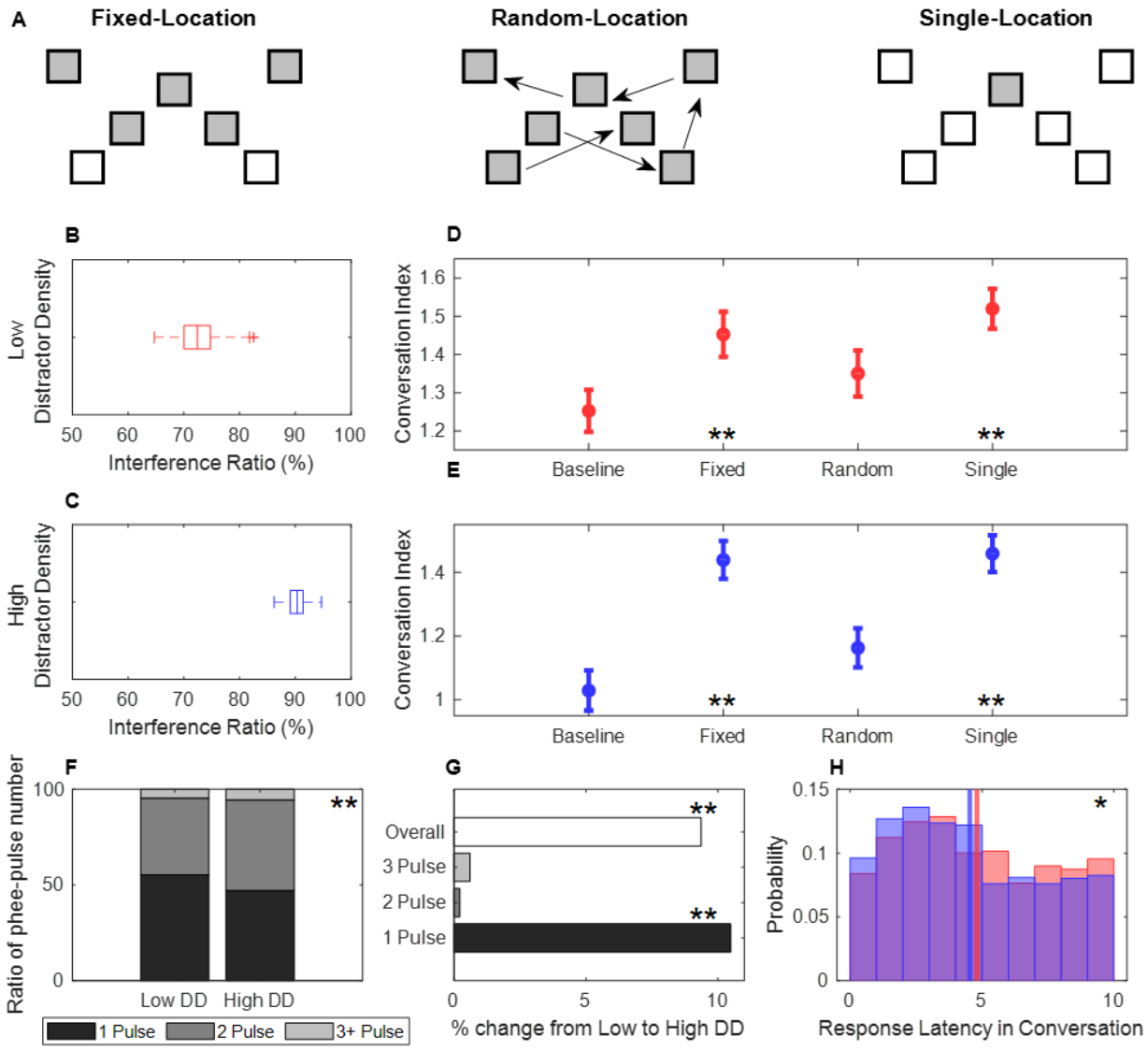

VM1
VM2
VM3
VM4
Target
Combined
Subject
0 10 20 30 40 50 60 70 80 90 100 110
Time (sec)

Figure 3.2: Experiment 1 Results.A schematic drawing of the spatial configuration of the seven speakers used in these three test conditions: Fixed-Location, Random-Location, and Single-Location. Grey shading indicates which speakers broadcast phee calls for that condition. Arrows in the Random-Location condition indicate the fact that the speaker location from which each VM phee was broadcast was randomized for each stimulus presentation across the seven-speakers. (B,C) Plots the Interference Ratio as measured by portion of Target VM calls that overlapped temporally with Distractor VM calls. Low Distractor Density (Low DD) is shown in red (B), while High Distractor Density (High DD) is shown in blue (C). (D, E) Plots the Mean Conversation index [95% CI] for Baseline, Fixed-Location, Random-Location, and Single-Location test conditions. ** Significant difference of condition from Baseline, $p < 0.0001$. (D) Plots Conversation Index for the Low Distractor Density condition, while (E) plots the High Distractor Density condition. (F) Stacked bar graph showing the distribution phee calls produced by subjects that comprised 1-Pulse (black), 2-Pulses (dark-grey) and 3 or more pulses (light-grey) in both the Low DD and High DD environments. ** Significant difference between distributions, $p < 0.0001$ (G) The change in duration of all calls, and sub-groups of phee-pulse calls from Low to High DD is shown as percent change. ** Significant difference for that category, $p < 0.0001$. (H) Histogram plots subjects' latency to respond to the Target VM in conversations in both Low DD (red) and High DD (blue) conditions. The median value is shown as a vertical red bar – Low DD – and blue bar – High DD. * Significant difference between distributions, $p < 0.05$.


A
Fixed-Location
Random-Location
Single-Location
B
Low Distractor Density
50 60 70 80 90 100
Interference Ratio (%)
C
High Distractor Density
50 60 70 80 90 100
Interference Ratio (%)
D
Conversation Index
1.6
1.5
1.4
1.3
1.2
**
**
Baseline
Fixed
Random
Single
E
Conversation Index
1.4
1.2
1
**
**
Baseline
Fixed
Random
Single
F
Ratio of phee-pulse number
100
50
0
**
Low DD
High DD
1 Pulse
2 Pulse
3+ Pulse
G
Overall
3 Pulse
2 Pulse
1 Pulse
**
**
0
5
10
% change from Low to High DD
H
Probability
0.15
0.1
0.05
0
*
0
5
10
Response Latency in Conversation

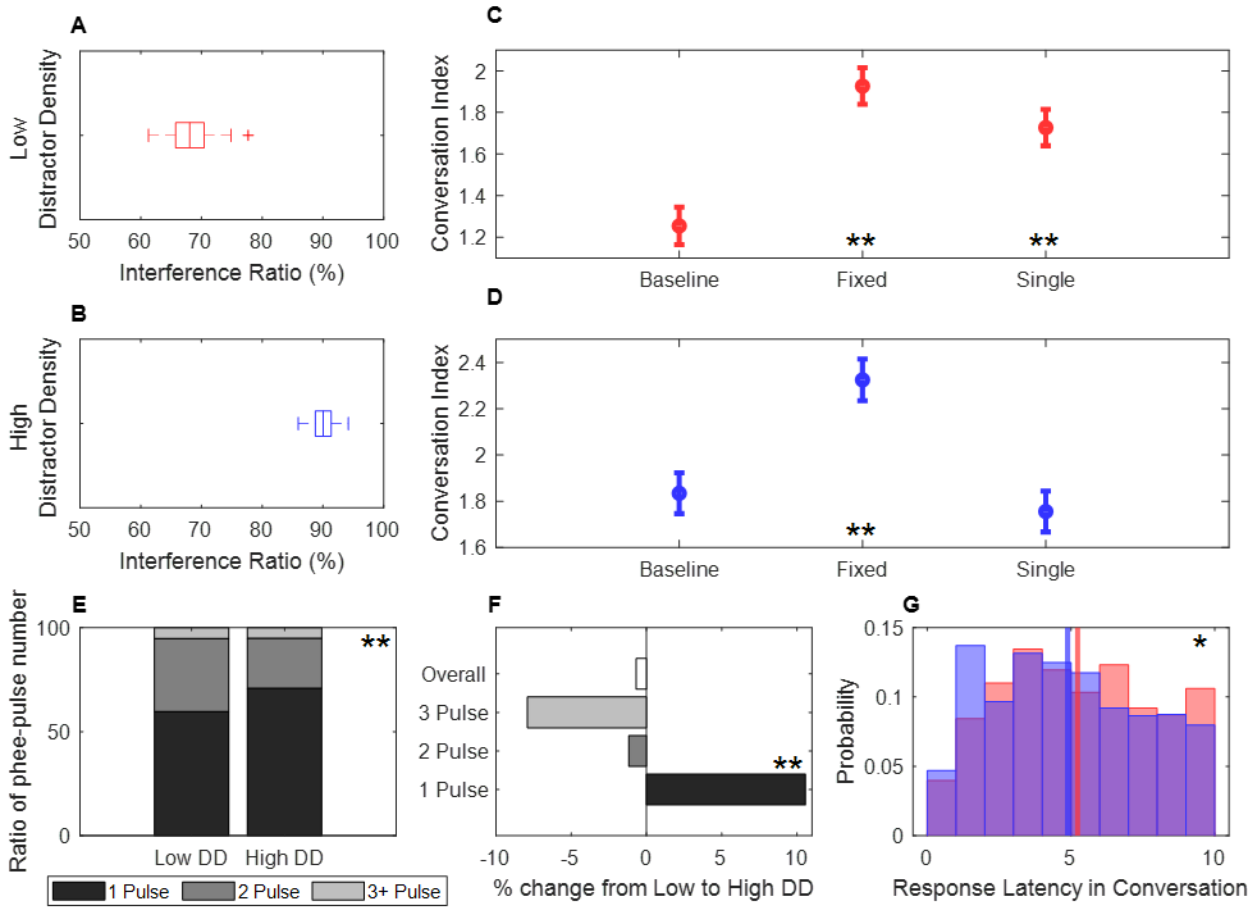
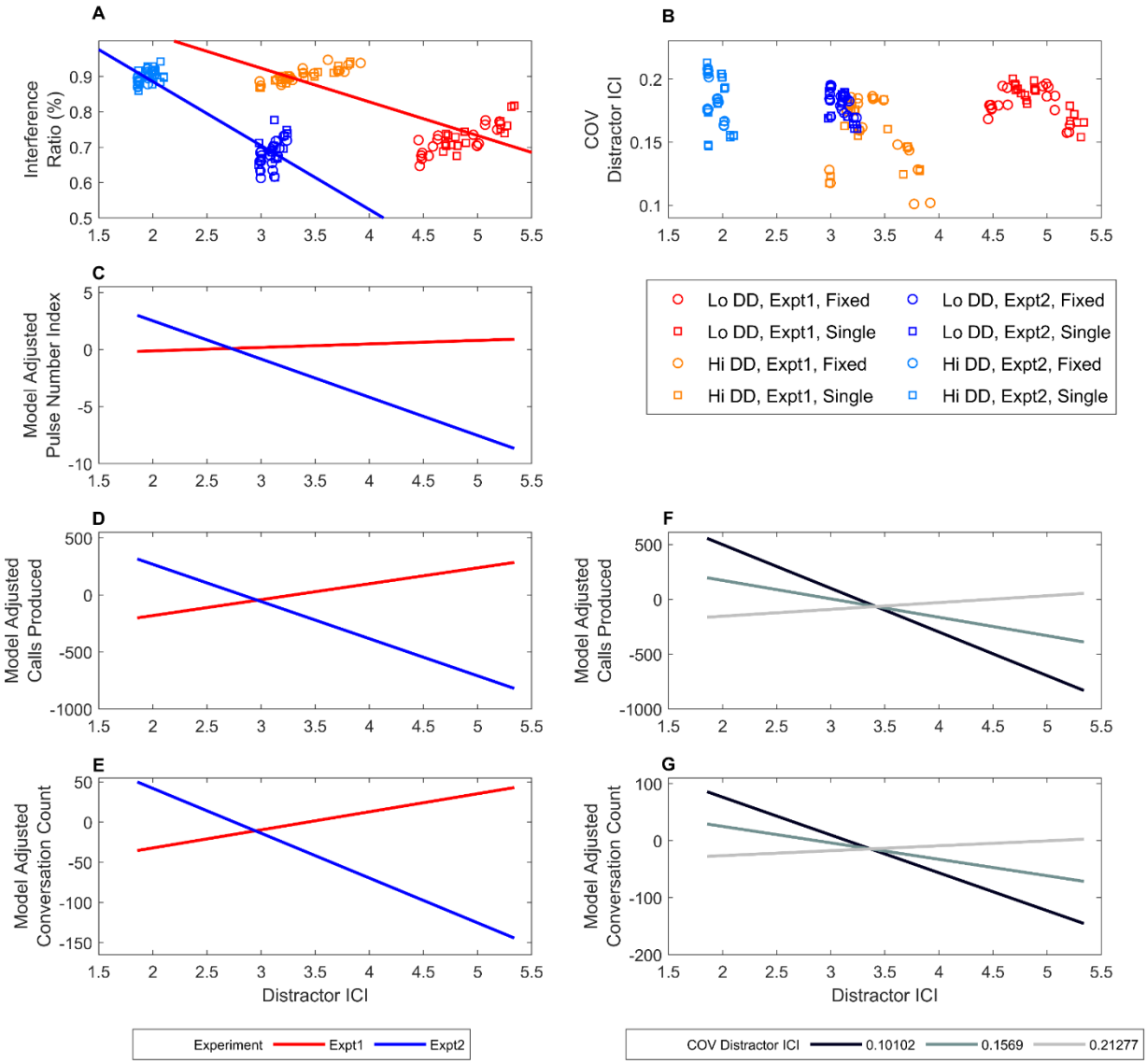


Figure 3.3: Experiment 2 Results. (A, B) Plots the Interference Ratio as measured by portion of Target VM calls that overlapped temporally with Distractor VM calls. Low Distractor Density (Low DD) is shown in red (A), while High Distractor Density (High DD) is shown in blue (B). (C, D) Plots the Mean Conversation index [95% CI] for Baseline, Fixed-Location and Single-Location test conditions. ** Significant difference of condition from Baseline, $p < 0.0001$. (C) Plots Conversation Index for the Low Distractor Density condition in red, while (D) plots the High Distractor Density condition in blue. (E) Stacked bar graph showing the distribution phee calls produced by subjects that comprised 1-Pulse (black), 2-Pulses (dark-grey) and 3 or more pulses (light-grey) in both the Low DD and High DD environments. ** Significant difference between distributions, $p < 0.0001$ (F) The change in duration of the phee calls comprising 1, 2, 3 and Overall duration is shown as percent change from Low DD to High DD conditions. ** Significant difference for that category, $p < 0.0001$ (G) Histogram plots subjects' latency to respond to the Target VM in conversations in both Low DD (red) and High DD (blue) conditions. The median value is shown as a vertical red bar – Low DD – and blue bar – High DD. * Significant difference for that category, $p < 0.01$.

Figure 3.4: Linear Model Outcome (A) Scatter plot displaying Interference Ratio for the Distractor ICI measured during in each test session. Lines represent the least-squares fit for each Experiment. (B) Plots the COV Distractor ICI for the Distractor ICI measured during in each test session. Figure legend for (A & B) is shown below (B). (C-E) Significant interactive effects of Distractor ICI with different metrics of vocal behavior revealed by the linear model are shown. Results of the model from Experiment 1: 2-pulse VM phee calls (red line) and Experiment 2: 1-pulse VM phee calls are shown (blue line). The adjusted response value accounts for the average values of all other terms except Distractor ICI x Experiment within the linear model. (C) Plots Distractor ICI by the adjusted response variable of Pulse Number Index. Pulse Number Index refers to the difference over sum of the portion of 1-pulse subject calls to 2-pulse subject calls. The more positive a value the higher the portion of 1-pulse phee calls subjects produced, while more negative values indicate a bias towards subjects producing 2-pulse phee calls. (D) Plots the relationship between the model adjusted calls produced by subjects by the Distractor ICI. (E) Plots the relationship between model adjusted Conversation Count (i.e. the number of at least two calls produced by a subject in succession with the target) by the Distractor ICI. (F,G) Distractor ICI x COV Distractor ICI term is plotted against its effect on the number Calls Subjects Produced (F) and Conversation Count (G). COV values plotted include minimum (light grey), maximum (dark grey), and the average of the two (mid-grey).


A
Interference Ratio (%)
B
COV Distractor ICI
C
Model Adjusted Pulse Number Index
Lo DD, Expt1, Fixed
Lo DD, Expt1, Single
Hi DD, Expt1, Fixed
Hi DD, Expt1, Single
Lo DD, Expt2, Fixed
Lo DD, Expt2, Single
Hi DD, Expt2, Fixed
Hi DD, Expt2, Single
D
Model Adjusted Calls Produced
F
Model Adjusted Calls Produced
E
Model Adjusted Conversation Count
Distractor ICI
G
Model Adjusted Conversation Count
Distractor ICI
Experiment
Expt1
Expt2
COV Distractor ICI
0.10102
0.1569
0.21277

3.8 References

Aubin T, Jouventin P (1998) Cocktail-party effect in king penguin colonies. Proceedings of the Royal Society of London Series B-Biological Sciences 265:1665-1673.

Bee MA (2015) Treefrogs as animal models for research on auditory scene analysis and the cocktail party problem. International Journal of Psychophysiology 95:216-237.

Bee MA, Micheyl C (2008) The cocktail party problem: What is it? How can it be solved? And why should animal behaviorists study it? J Comp Psych 122:235-251.

Bendor DA, Wang X (2008) Neural response properties of the primary, rostral and rostrotemporal core fields in the auditory cortex of marmoset monkeys. J Neurophys 100:888-906.

Bey C, McAdams S (2002) Schema-based processing in auditory scene analysis. Perception & Psychophysics 64:844-854.

Bregman AS (1994) Auditory Scene Analysis: The Perceptual Organization of Sound. Cambridge, MA: MIT Press.

Bronkhorst AW (2015) The cocktail-party problem revisited: early processing and selection of multi-talker speech. Attention, Perception, & Psychophysics 77:1465-1487.

Brungart DS, Simpson BD (2002) The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal. J Acoust Soc Am 112:664-676.

Brungart DS, Simpson BD (2007) Cocktail party listening in a dynamic multitalker environment. Perception & Psychophysics 69:79-91.

Cherry EC (1953) Some experiments on the recognition of speech, with one and with two ears. J Acoust Soc Am 25:975-979.

Chow C, Mitchell J, Miller CT (2015) Vocal turn-taking in a nonhuman primate is learned during ontogeny. Proceedings of the Royal Society, B 282:210150069.

Darwin CJ (1997) Auditory grouping. Trends in Cognitive Sciences 1:327-333.

Darwin CJ, Hukin RW (2000) Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention. The Journal of the Acoustical Society of America 108:335-342.

de la Mothe LA, Blumell S, Kajikawa Y, Hackett TA (2006) Cortical connections of the auditory cortex in marmoset monkeys: core and medial belt regions. J Comp Neurol 496:27-71.

Eliades SJ, Miller CT (2017) Marmoset vocal communication: Neurobiology and behavior. Developmental Neurobiology 77:286-299.

Hackett TA (2009) The evolution of primate and human auditory system. In: Evolutionary Neuroscience (Kaas JH, ed), pp 893-903. San Diego, CA: Academic Press.

Hill KT, Miller LM (2009) Auditory Attentional Control and Selection during Cocktail Party Listening. Cerebral Cortex 20:583-590.

Johnsrude IS, Mackey A, Hakyemez H, Alexander E, Trang HP, Carlyon RP (2013) Swinging at a Cocktail Party:Voice Familiarity Aids Speech Perception in the Presence of a Competing Voice. Psych Sci 24:1995-2004.

Jones GL, Litovsky RY (2011) A cocktail party model of spatial release from masking by both noise and speech interferers. The Journal of the Acoustical Society of America 130:1463-1474.

Kaas JH (2010) Sensory and motor systems in primates. In: Primate Neuroethology (Platt M, Ghazanfar AA, eds), pp 177-200. New York, NY: Oxford University Press.

Kaas JH, Hackett TA (1998) Subdivisions of auditory cortex and levels of processing in primates. Audiology and Neuro-Otology 3:73-85.

Kaas JH, Hackett TA (2000) Subdivisions of auditory cortex and procesing streams in primates. PNAS 97:11793-11799.

Kidd G, Arbogast TL, Mason CR, Gallun FJ (2005) The advantage of knowing where to listen. The Journal of the Acoustical Society of America 118:3804-3815.

Lee N, Ward JL, Vélez A, Micheyl C, Bee MA (2017) Frogs Exploit Statistical Regularities in Noisy Acoustic Scenes to Solve Cocktail-Party-like Problems. Current Biology 27:743-750.

Litovsky RY (2005) Speech intelligibility and spatial release from masking in young children. The Journal of the Acoustical Society of America 117:3091-3099.

Mcdermott JH (2009) The cocktail party problem. Current Biology 19:R1024-R1027.

Micheyl C, Tian B, Carlyon RP, Rauschecker JP (2005) Perceptual organization of tone sequences in the auditory cortex of awake Macaques. Neuron 48:139-148.

Micheyl C, Carlyon RP, Shtyrov Y, Hauk O, Dodson T, Pullvermuller F (2003) The neurophysiological basis of the auditory continuity illuion: A mismatch negativity study. J Cog Neurosci 15:747-758.

Miller CT, Wang X (2006) Sensory-motor interactions modulate a primate vocal behavior: antiphonal calling in common marmosets. Journal of Comparative Physiology A 192:27-38.

Miller CT, Thomas AW (2012) Individual recognition during bouts of antiphonal calling in common marmosets. Journal of Comparative Physiology A 198:337-346.

Miller CT, Dibble E, Hauser MD (2001) Amodal completion of acoustic signals by a nonhuman primate. Nature Neuroscience 4:783-784.

Miller CT, Iguina C, Hauser MD (2005) Processing vocal signals for recognition during antiphonal calling. Anim Behav 69:1387-1398.

Miller CT, Eliades SJ, Wang X (2009a) Motor-planning for vocal production in common marmosets Anim Behav 78:1195-1203.

Miller CT, Mandel K, Wang X (2010) The communicative content of the common marmoset phee call during antiphonal calling. . Am J Primatol 72:974-980.

Miller CT, Beck K, Meade B, Wang X (2009b) Antiphonal call timing in marmosets is behaviorally significant: Interactive playback experiments. . Journal of Comparative Physiology A 195:783-789.

Miller CT, Freiwald W, Leopold DA, Mitchell JF, Silva AC, Wang X (2016) Marmosets: A Neuroscientific Model of Human Social Behavior. Neuron 90:219-233.

Mitani J, Stuht J (1998) The evolution of nonhuman primate loud calls: acoustic adaptation for long-distance transmission. Primates 39:171-182.

Morrill R, Thomas AW, Schiel N, Souto A, Miller CT (2013) The effect of habitat acousics on common marmoset vocal signal transmission. Am J Primatol:904-916.

Pastore MT, Yost WA (2017) Spatial Release from Masking with a Moving Target. Frontiers in Psychology 8.

Petkov CI, O'Connor KN, Sutter ML (2003) Illusory sound perception in macaque monkeys. J Neurosci 23:9155-9161.

Pomberger T, Löschner J, Hage SR Compensatory mechanisms affect sensorimotor integration during ongoing vocal motor acts in marmoset monkeys. European Journal of Neuroscience n/a.

Pomberger T, Risueno-Segovia C, Gultekin YB, Dohmen D, Hage SR (2019) Cognitive control of complex motor behavior in marmoset monkeys. Nature Communications 10:3796.

Pressnitzer D, Sayles M, Micheyl C, Winter IM (2008) Perceptual Organization of Sound Begins in the Auditory Periphery. Current Biology 18:1124-1128.

Roy S, Miller CT, Gottsch D, Wang X (2011) Vocal control by the common marmoset in the presence of interfering noise. J Exp Biol 214:3619-3629.

Shinn-Cunningham BG (2008) Object-based auditory and visual attention. Trends in Cognitive Sciences 12:182-186.

Song X, Osmanski MS, Guo Y, Wang X (2016) Complex pitch perception mechanisms are shared by humans and a New World monkey. PNAS 113:781-786.

Toarmino C, Wong L, Miller CT (2017) Audience affects decision-making in a marmoset communication network. Biology Letters 13:20160934.

Wang X (2013) The harmonic organization of auditory cortex. Frontiers in Neuroscience 7:2013.00114.

Wang X, Walker KMM (2012) Neural mechanisms for the abstraction of pitch information in auditory cortex. J Neurosci 32:13339-13342.

Waser PM, Waser MS (1977) Experimental studies of primate vocalization - specializations for long-distance propagation. Zeit Tierpsychol 43:239-263.

Waser PM, Brown CH (1986) Habitat acoustics and primate communication. Am J Primatol 10:135-154.

Woods KJP, McDermott JH (2018) Schema learning for the cocktail party problem. PNAS 115:E3313.

Zhao L, Rad BB, Wang X (2019) Long-lasting vocal plasticity in adult marmoset monkeys. Proceedings of the Royal Society B: Biological Sciences 286:20190817.