PNAS RESEARCH ARTICLE | NEUROSCIENCE

Anterior cingulate cortex neurons parse speakers in a natural cocktail party

Arthur Lefevre[a,b,1], Vikram Pal Singh[a], Timothy J. Tyree[a], Jingwen Li[a], Pia Giraudet[b], Jean-René Duhamel[b], and Cory T. Miller[a,c]

Affiliations are included on p. 9.

Edited by Lizabeth Romanski, University of Rochester, Rochester, NY; received January 16, 2026; accepted October 20, 2026 by Editorial Board Member Huda Akil

The cocktail party problem—parsing relevant voices from competing speakers—has traditionally been studied as a sensory challenge for auditory cortex. Yet in natural communication, identifying speakers and coordinating interaction are inherently social processes. Here we recorded single neurons from anterior cingulate cortex (ACC, area 24) in freely moving marmosets engaged in spontaneous vocal exchanges within a noisy colony environment. ACC neurons selectively encoded vocalizations from a social partner versus other callers in the scene, effectively parsing speakers in the acoustic mixture. This selectivity was preserved even when vocalizations overlapped in time, indicating robustness to background interference. Apart from perception, ACC population activity represented vocal production, call type, and whether calls occurred within vocal turn-taking exchanges. Behaviorally, marmosets adjusted vocal timing based on the surrounding acoustic scene, and these interaction contexts were reliably reflected in ACC activity. Together, these findings suggest that neurons in ACC play a role in parsing conspecific speakers during communication in a natural "cocktail party," linking auditory scene analysis with social monitoring and coordinated interactions.

anterior cingulate cortex | cocktail party | vocal communication | natural behavior | social behavior

Significance

In noisy social settings, such as human gatherings or animal colonies, identifying and responding to relevant voices amid competing sounds—the "cocktail party problem"—is a fundamental challenge. This study reveals that neurons in the anterior cingulate cortex (ACC) of freely interacting marmosets selectively encode vocalizations from social partners, even when calls overlap in time. Beyond perception, ACC activity reflects vocal production, call type, and the structure of turn-taking exchanges. These findings suggest that the ACC is involved in integrating auditory scene analysis with social monitoring and coordinated action, providing insight into the neural mechanisms underlying natural communication. The results may also inform our understanding of communication disorders, such as autism, where social interaction and auditory processing are impaired.

The Cocktail Party Problem (CPP)—the perceptual phenomenon by which we extract meaningful sounds from noise—is arguably the most persistent real-world challenge for acoustic communication systems, such as speech (1). Despite its fundamental importance and decades of research, the neural mechanisms underlying the CPP remain elusive, particularly in human and nonhuman primates (2). This gap reflects at least three limitations in prior research.

First, most neurobiological research related to the CPP has been performed in highly controlled laboratory settings that minimize uncertainty, competition, and social contingency (1–3). Under these conditions, resolving the CPP is reduced to a sensory discrimination problem, placing limited demands on neural systems involved in monitoring and adaptively responding to conspecifics. Notably, this context contrasts considerably with the acoustically dynamic, naturalistic environments where the CPP was first described (3) and naturally arises. Second, CPP research has focused primarily on neocortical regions associated with human speech processing (2, 4–6), reflecting a speech-centric and phylogenetically narrow view of the problem. This emphasis obscures the fact that the CPP predates language and the expansion of these cortical circuits in primate evolution, suggesting that more conserved neural systems likely contribute to its resolution. Third, beyond the choice of experimental setting, the CPP has been conceptualized primarily as an auditory computation. In real-world social interactions, however, the problem is not simply identifying a sound source, but determining which signal is relevant, when to respond, and how that response fits within an ongoing exchange with conspecifics—demands that require tracking social context and interaction history rather than acoustic features alone. Such functions are not central features of the canonical auditory pathways (7), but are hallmarks of brain regions involved in cognitive control and social monitoring.

Together, these gaps motivate consideration of brain regions traditionally outside speech and auditory processing networks. One such candidate is the Anterior Cingulate Cortex (ACC), which has been largely ignored in studies of CPP despite its established role in vocal production (8), vocal perception (9), and, critically, social monitoring during interactions with conspecifics (10). Because resolving the CPP enables individuals to communicate with conspecifics in complex natural environments (11, 12), we posited that uncovering its neural basis likely requires its study under similarly dynamic conditions (13).

Author contributions: A.L., J.-R.D., and C.T.M. designed research; A.L. and V.P.S. performed research; A.L., T.J.T., J.L., and P.G. analyzed data; and A.L. and C.T.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. L.R. is a guest editor invited by the Editorial Board.

Copyright © 2026 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]To whom correspondence may be addressed. Email: arthur.lefevre@isc.cnrs.fr.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2601809123/-/DCSupplemental.

Published November 24, 2026.

Downloaded from https://www.pnas.org by UC San Diego on November 24, 2026 from IP address 64.98.200.126.

To address this issue, we leveraged the rich vocal communication system of marmoset monkeys (14)—highly vocal New World primates (15, 16). Marmosets engage in a range of context-dependent vocal exchanges, including turn taking with different call types (12), and produce vocalizations that are modulated by social dynamics (17–19). Moreover, these primates thrive in acoustically cluttered environments, making them an ideal species for studying the neurobiology of the CPP under real-world conditions.

We hypothesized that the ACC Area 24 will exhibit at least the three following mechanisms integral to resolving the CPP in natural communication scenes: i) selectivity for callers' identity and calls' type, ii) respond invariantly to vocalizations regardless of background sounds and iii) encode structure of active vocal turn taking exchanges. To test these hypotheses, we recorded the activity of individual ACC neurons in freely-moving, naturally behaving marmosets in their home cage within the acoustically rich colony room (Fig. 1*A*), enabling detailed analysis of neural

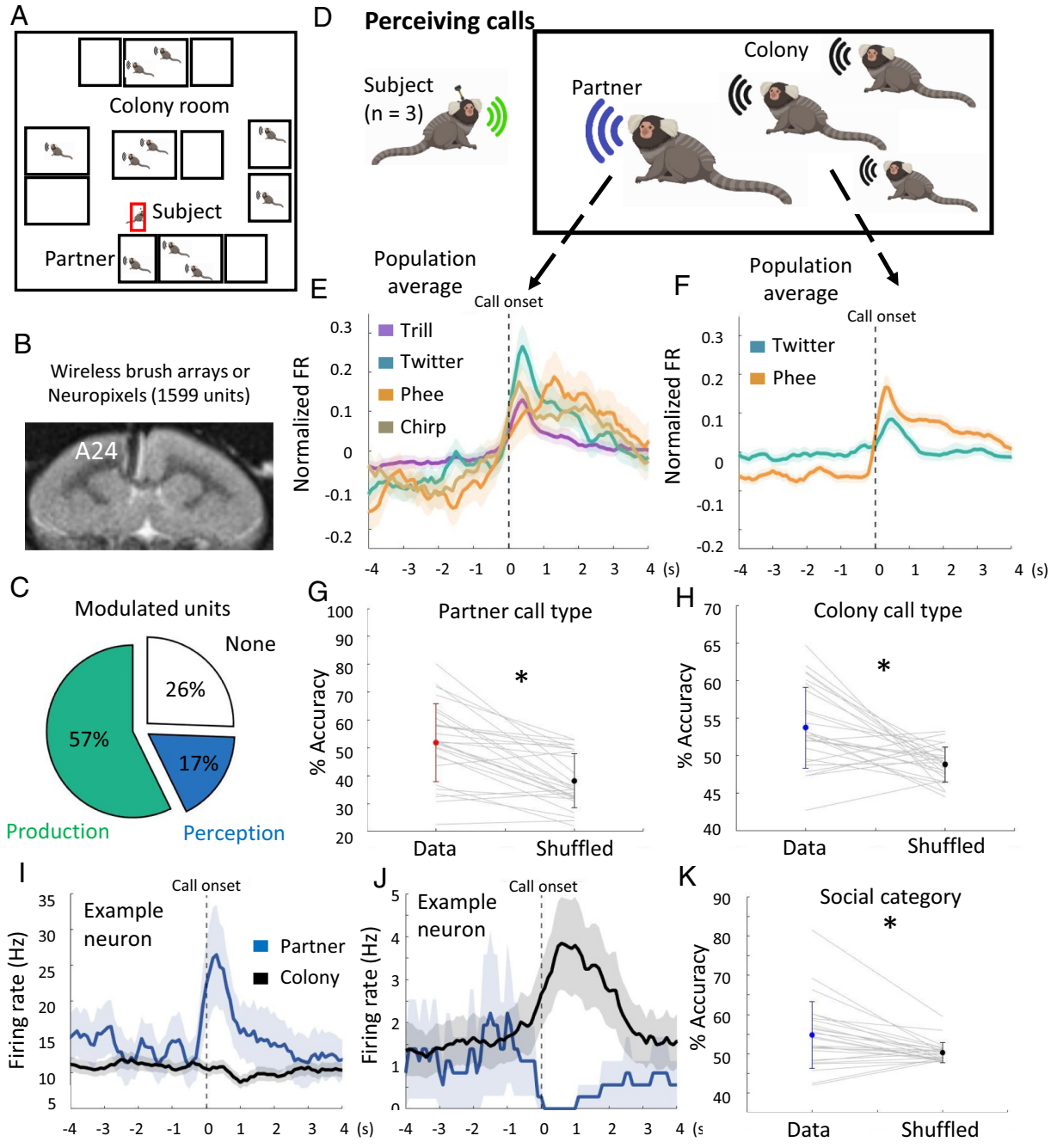


Fig. 1. Segregated ACC neuronal populations encode vocalizations perceived from the partner and the colony. (*A*) Schema of the experimental paradigm, recordings took place directly in the home cage and transfer box in the colony room. (*B*) MRI coronal plane (AP = +12 mm) at which probes (brush arrays or neuropixels) were implanted. (*C*) Percentage of neurons responding to vocalization production and perception. (*D*) Vocalizations produced by the partner or other marmosets in the colony (vs. produced ones) were analyzed separately. (*E*) Normalized average increase from neurons with significantly upregulated activity during partner call perception. (*F*) Normalized average increase from neurons with significantly upregulated activity during colony call perception. (*G* and *H*) Percentage accuracy of the neuronal multiclass decoder for partner (*G*) and colony (*H*) perceived call type for real data and data with shuffled labels. (*I*) Example neuron responding specifically to twitter calls from the partner but not from the colony. (*J*) Example neuron responding specifically to phee calls from the colony but not from the partner. (*K*) Percentage accuracy of the neuronal SVM decoder for identity (partner vs. colony) for real data and data with shuffled labels. *$P < 0.01$, error bars indicate SEM, (SE to the mean).

Downloaded from https://www.pnas.org by UC San Diego on November 24, 2026 from IP address 64.98.200.126.

responses during natural conversations with their cage-mate partner or with members of the colony amid ongoing background vocalizations from other conspecifics.

Results

Segregated ACC Neuronal Populations Encode Vocalizations Perceived From the Partner and the Colony. All experiments were performed in the noisy, colony room housing ~60 animals where subjects are housed. The subject (2 females, 1 male) was placed in a small transparent box in front of their home cage that contained the partner (Fig. 1*A*). Using wearable microphones and an automated labeling algorithm (20), analyses distinguished whether each call was produced by the subject, their partner, or the colony. We recorded ACC area 24 wirelessly using micro wire brush arrays or Neuropixels probes (Fig. 1*B* and *SI Appendix*, Fig. S1), yielding 1,599 neurons and 38,559 vocalizations across 46 sessions (Fig. 1*B* and *SI Appendix*, Fig. S2). Use of chronic arrays can result in sampling of the same neurons on successive days, however, our moveable brush arrays and previous literature on Neuropixels probes indicate that over 50% of our units were recorded only once (*Material and Methods*).

Similar to previous work (8, 21–24), we observed that ACC neurons were predominantly motor modulated, though a population were responsive to hearing the calls of conspecifics (i.e., vocalization perception) (Fig. 1*C*). While ACC exhibits sensory responses to passive presentations of vocalizations (9, 25), whether this reflects a general response to conspecific calls or neurons in this substrate will display a selectivity to different call types necessary to resolve the CPP during natural vocal communication contexts was not clear. To test this, we first analyzed ACC neuron activity in response to hearing their partner's vocalizations and those produced by other conspecifics in the colony (Fig. 1*D*). Analyses revealed that neurons in area 24 responded robustly to both partner calls (Wilcoxon signed rank tests, $P < 0.05$, n = 167, Fig. 1*E* and *SI Appendix*, Fig. S3*C*) and to colony calls (Wilcoxon signed rank tests, $P < 0.05$, n = 77, Fig. 1*F*). The majority of significantly modulated neurons responded to only a single call type (86.8%, *SI Appendix*, Fig. S3*D*). Given this degree of call type specificity in the activity of individual ACC neurons, we investigated whether call types were robustly encoded at the population level. A multiclass decoder, trained on neural activity of all neurons from 0 s to +3 s from call onset, revealed that individual call types from both the partner (Wilcoxon signed rank test, z = 4.16, $P < 0.001$) (Fig. 1*G* and *SI Appendix*, Fig. S4) and the colony (Wilcoxon signed rank test, z = 3.19, $P < 0.001$) (Fig. 1*H* and *SI Appendix*, Fig. S4) could be significantly decoded from ACC population activity.

We next examined whether ACC area 24 neurons separately encode different types of callers in the scene, namely partner calls and those produced by any other marmosets in the colony. In other words, do these neurons simply encode call type or a combination of call type and social categories? We observed two distinct populations of neurons, one that responded only when hearing partner calls (Fig. 1*I* and *SI Appendix*, Fig. S5*A*) and the second that responded only to hearing any other caller in the colony (Fig. 1*J* and *SI Appendix*, Fig. S5*B*). This suggests that ACC neurons parse conspecific calls into two categories—partner and others (i.e., colony)—and encode both. While these neural populations were almost entirely segregated; 2 neurons (out of 244) responded to both partner and colony calls, suggesting the existence of a minority of neurons with general auditory responses. Moreover, the selectivity was not biased to partner calls, with similar proportions of neurons encoding both categories of callers within the same ACC substrate (5.3% of neurons encoding partner phees or twitters, 4.8% encoding colony phees or twitters). At the very least, these neurons suggest that figure/ground separation of highly meaningful signals (1, 3, 26) is encoded in ACC area 24, a foundational mechanism for resolving the CPP in natural acoustic scenes.

To further validate this idea at the population level, we used neural population activity to decode whether a call had been produced by the partner or the colony using a support vector machine (SVM) classifier. Analyses revealed significant decoding of these social categories (Wilcoxon signed rank test, z = 2.91, $P = 0.0013$) (Fig. 1*K*). Importantly, additional analyses revealed that this segregation was not caused by differences in spatial properties (location or distance, *SI Appendix*, Fig. S6) or sound intensity (Fig. 2 *E* and *F* and *SI Appendix*, Fig. S7).

ACC Neurons Respond Invariantly to Calls With and Without Overlapping Vocalizations. To fully resolve the CPP, neural mechanisms that parse meaningful signals (such as partner's voice) from acoustically overlapping background sounds are necessary (1, 3), particularly in the presence of similar vocalizations produced by conspecifics. To more closely examine how ACC responses were affected by background sounds, we investigated how neurons responded to vocalizations in the two following contexts (26): "Simultaneous calls" [i.e., calls that occurred concurrently with a colony call (Fig. 2*A*)] and "'Sequential calls" [i.e., calls that had no acoustic overlap with other vocalizations (Fig. 2*B*) (26)]. Analyses focused on twitter calls because it had the largest population of perceptual neurons selective to partner calls (n = 54 neurons), thereby allowing comparison in both contexts.

Analyses revealed that although neurons were strongly driven to partner twitter calls (Wilcoxon signed rank tests, $P < 0.05$), there was no statistical difference in responses between the sequential and simultaneous contexts both at the single cell (all P values > 0.05, Fig. 2*C*) and at the population levels ($P = 0.49$, Fig. 2*D*), indicating that background sounds do not affect the responses of these neurons. In other words, single neurons in marmoset ACC area 24 selectively parse the meaningful signal—vocalizations produced by the partner—from overlapping background sounds (e.g., all other conspecific vocalizations), thereby exhibiting a key mechanism necessary to resolve the CPP. Following this logic, populations of neurons encoding partner calls and colony calls were almost entirely distinct (Fig. 1 *J* and *K*). Although partner and colony twitters and phees were of significantly different sound intensity (dB, t test $P < 0.001$), further analyses controlling for this factor showed that it did not impact this key result (Fig. 2 *E* and *F* and *SI Appendix*, Fig. S7). This suggests that attentional mechanisms in ACC (27–29) may enable neurons in this substrate to encode distinct calls produced by conspecifics that are temporally simultaneous.

Marmosets Monitor Social Scenes to Optimize Vocal Turn Taking in a "Cocktail Party". Resolving the CPP is essential for vocal communication in both human and nonhuman animals, where the ability to parse, produce, and respond to vocal signals amid background noise enables meaningful interactions. To better understand how marmosets communicate in such environments, we examined vocal turn taking with different vocalizations in the noisy colony environment, a natural behavior observed in many animals including humans and marmosets (12).

To quantify marmoset turn-taking in this context, we focused analyses on two distinct vocalization types—trills (14, 20) and twitters (14) (Fig. 3 *A* and *B*)—by aligning the partner and colony call rates on the subject's vocalizations (Fig. 3 *C* and *D*). Trills are

Downloaded from https://www.pnas.org by UC San Diego on November 24, 2026 from IP address 64.98.200.126.

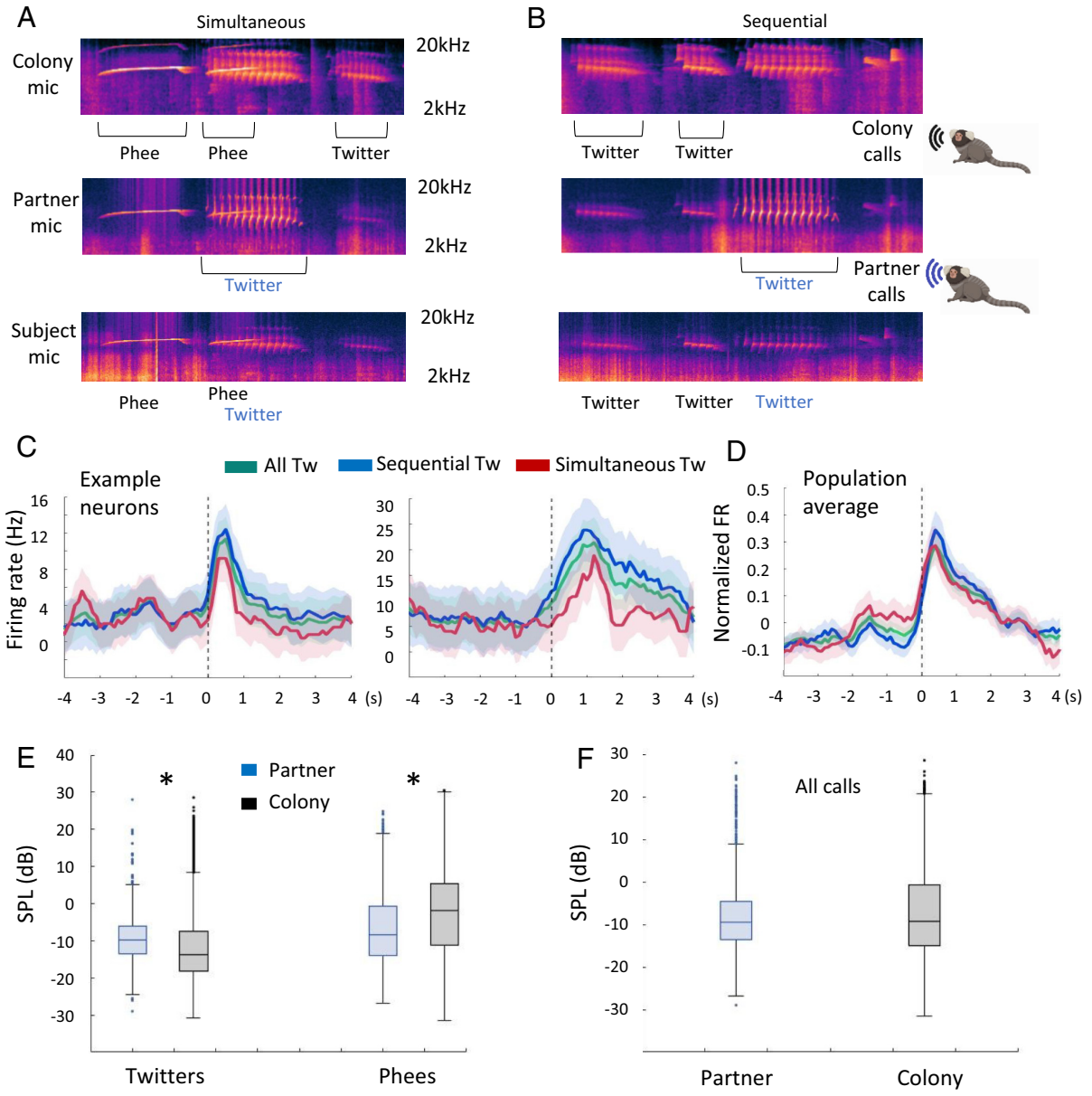


Fig. 2. ACC neurons respond invariantly to calls with and without overlapping vocalizations. (*A*) Example spectrograms from the colony and partner microphones showing the occurrence of a call from the partner (twitter) that is perceived simultaneously with a call from the colony (phee). (*B*) Example spectrograms from the colony and partner microphones showing the occurrence of a call from the partner (twitter) that is perceived sequentially to calls from the colony (twitters). (*C*) Example neurons responding to twitter calls from the partner that are unaffected by the presence (or absence) of simultaneous calls from the colony. (*D*) Normalized average increase from neurons with significantly upregulated activity during partner twitter call perception is unaffected by the presence (or absence) of simultaneous calls from the colony. (*E*) Sound Pressure Levels (SPL in dB, measured in the subject's microphone) for twitters and phees from the partner and colony marmosets. (*F*) Sound Pressure Levels for all phees and twitters after removing 15% of lowest colony twitters and 20% of loudest colony phees to equalize SPL ($P = 0.39$). *$P < 0.01$, error bars indicate SEM (*C* and *D*) or SD (*E* and *F*).

low-amplitude calls exchanged between affiliated conspecifics in close proximity (14, 20), whereas twitters are composed of multiple sharply rising notes, mid-distance contact calls, with less affiliative component (14, 30). This natural ethological difference between these call types during turn-taking provided an opportunity to investigate the behavioral strategies marmosets use to communicate in noisy acoustic scenes. We found that partners produced significantly more trills in close temporal proximity to the subject's trill onset (sliding Wilcoxon's signed rank tests, 0.1 to 1.4 s, $P < 0.001$, Fig. 3*C*) but no such relation was found for twitters between partners (Fig. 3*C*). Moreover, subjects produced the low amplitude trill calls when the colony twitter call rate was significantly lower than average (−1.5 to 0.6 s, $P < 0.001$, Fig. 3*D*), effectively avoiding noisy periods in the colony. Finally, we found increase in colony twitter rate before and after subjects' twitter meaning subjects were twittering in synchrony with the colony (1.5 to 2.5 s), while avoiding interrupting each other, as represented by lower colony twitter rate at subjects' twitter call onset (0.5 to 1.6 s, $P < 0.001$, Fig. 3*D*). This pattern suggests that marmosets monitor the colony's sound level and adopt context-dependent strategies for turn-taking with each call type to maximize respective communication signaling efficacy.

ACC Neurons Encode Active Turn Taking With the Partner and the Colony. Although the CPP has been considered mainly as a perceptual problem, it is integral to communication in noisy socio-acoustic environments. Thus, in the present paradigm, our subjects were not only perceiving calls but also producing them,

Downloaded from https://www.pnas.org by UC San Diego on November 24, 2026 from IP address 64.98.200.126.

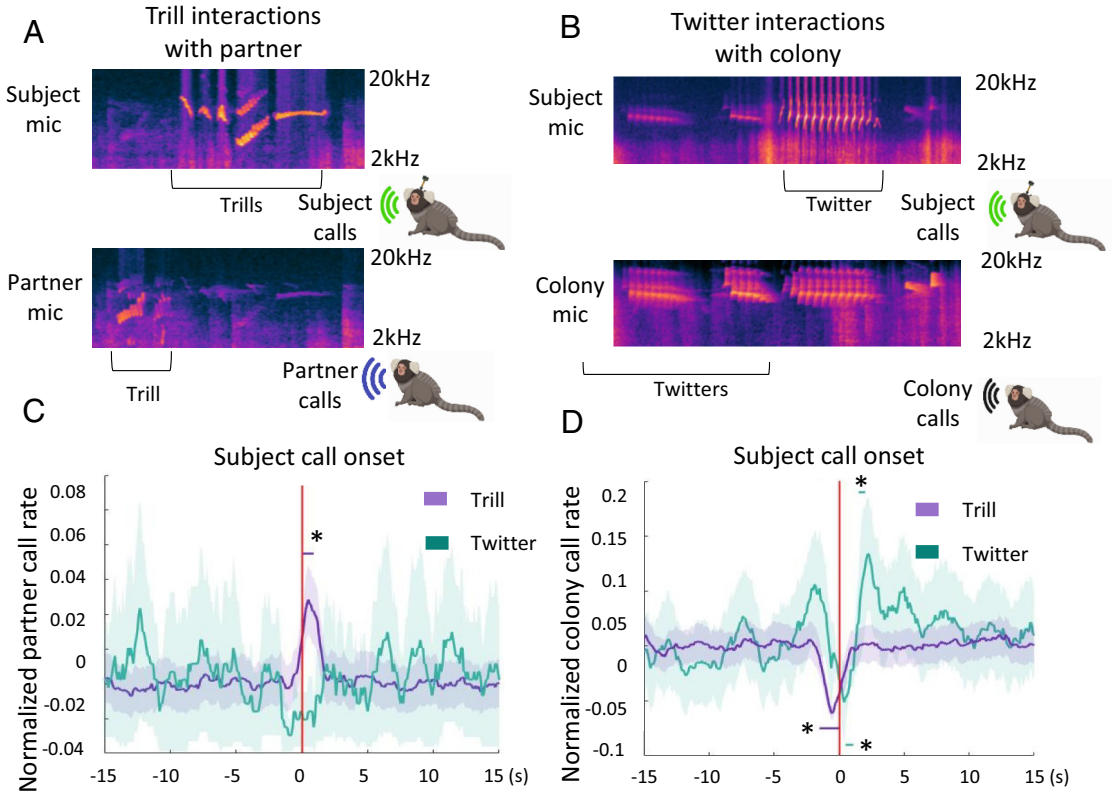


Fig. 3. Marmosets monitor social scenes to optimize vocal turn taking in a cocktail party. (*A*) Example spectrograms showing trill to trill conversations between the subject and its partner. (*B*) Example spectrograms showing twitter to twitter conversations between the subject and other monkeys in the colony. (*C*) Normalized partner call rate for trills (purple) and twitters (green) aligned on subject's trills and twitters onsets. (*D*) Colony call rate for twitters aligned on subject's trills (purple) and twitters (green). *$P < 0.01$

reflecting active communication. We analyzed neural activity during vocal production. Results showed that neurons exhibited a significant increased (n = 311/1,599, 19.4%) or decreased (n = 652/1,599, 40.8%) activity when the subject was producing a call (Wilcoxon signed rank tests, $P < 0.05$) (Fig. 4 *A–C* and *SI Appendix*, Fig. S3 *B–D*). Like call perception (Fig. 1*F*, G), the majority of neurons were highly selective to only one call type (83.3%, *SI Appendix*, Fig. S3*E*). However, a subset of neurons (18%) exhibited responses to multiple types of vocalizations (*SI Appendix*, Fig. S3*E*). Interestingly, we observed that few neurons encoded both production and perception of calls (n = 51, *SI Appendix*, Fig. S3*E*), but these units did not consistently encode the same call type across modalities (*SI Appendix*, Fig. S8*A* and Table S1).

We quantified whether the produced call type could be robustly decoded from neural population activity using a multiclass decoder trained on calls produced by subjects. Results showed that call types could be accurately classified based on ACC activity (Wilcoxon signed rank test, z = 4.54, $P < 0.001$) (Fig. 4*D* and *SI Appendix*, Fig. S9). This decoding was dependent on the number of neurons rather than on the ones significantly modulated by call production (*SI Appendix*, Fig. S9). Together, these data show that ACC area 24 neurons encode call production before, during, and after call onset with a high degree of selectivity, consistent with results in other motor systems (31). Furthermore, we tested if ACC represented call type across modalities (call production and perception). We trained a decoder on calls produced and tested classification accuracy on calls perceived, and vice versa. This analysis revealed that decoding accuracy was significantly better than chance (Wilcoxon signed rank tests, $P < 0.001$, *SI Appendix*, Fig. S8 *B* and *C*) suggesting some level sensory motor integration of call type in the neural population.

To further investigate whether ACC neurons are sensitive to the social context in which calls are produced, we employed a time sensitive population decoding analysis designed to identify periods with maximal differences in neural activity between different contexts [see *Material and Methods* and (32) for details]. Specifically, we applied this approach to test whether population activity could reliably classify whether a given call occurred as part of a structured turn-taking exchange or independent of these social interactions. Turn-taking was defined as an event in which subjects produced a call within 5 s (for partner trill calls) or 10 s (for colony twitter calls) of a conspecific producing the same call type (Fig. 3*A*, B).

Analyses revealed that the behavioral context – whether the vocalization occurred in a turn-taking exchange or not – could be accurately classified from ACC area 24 population activity. Trill calls produced (accuracy: 86.9%) or perceived (accuracy: 82.7%) (Fig. 4*E* and *SI Appendix*, Fig. S10*C*) were accurately classified as independent or part of a turn-taking bout with the partner, according to one-sided Mann–Whitney's U-tests ($z = 2.48 \times 10^5$, $P < 0.001$, $N_{calls} = 596$ produced, $N_{calls} = 776$ perceived). Likewise, twitter calls produced (accuracy: 68.3%) or perceived (accuracy: 96.0%) (Fig. 4*F* and *SI Appendix*, Fig. S10*D*) were classified with significant statistical accuracy according to a one-sided Mann–Whitney's *U* test (produced: $z = 3.1 \times 10^4$, $P = 0.011$, $N_{calls} = 236$; perceived: $z = 8.5 \times 10^4$, $P < 0.001$, $N_{calls} = 314$).

The CPP, and the conversational exchanges evidenced above, certainly require strong attentional capacities (1). Thus the ACC, a substrate known to be involved in social attention (28), may likewise be involved in the attentional processes that govern the vocal behaviors examined here. Although our freely moving natural paradigm was not designed for this question, we took two approaches using subjects' vocal behavior as a proxy of attention to investigate this issue. First, we conjectured that producing a

Downloaded from https://www.pnas.org by UC San Diego on November 24, 2026 from IP address 64.98.200.126.

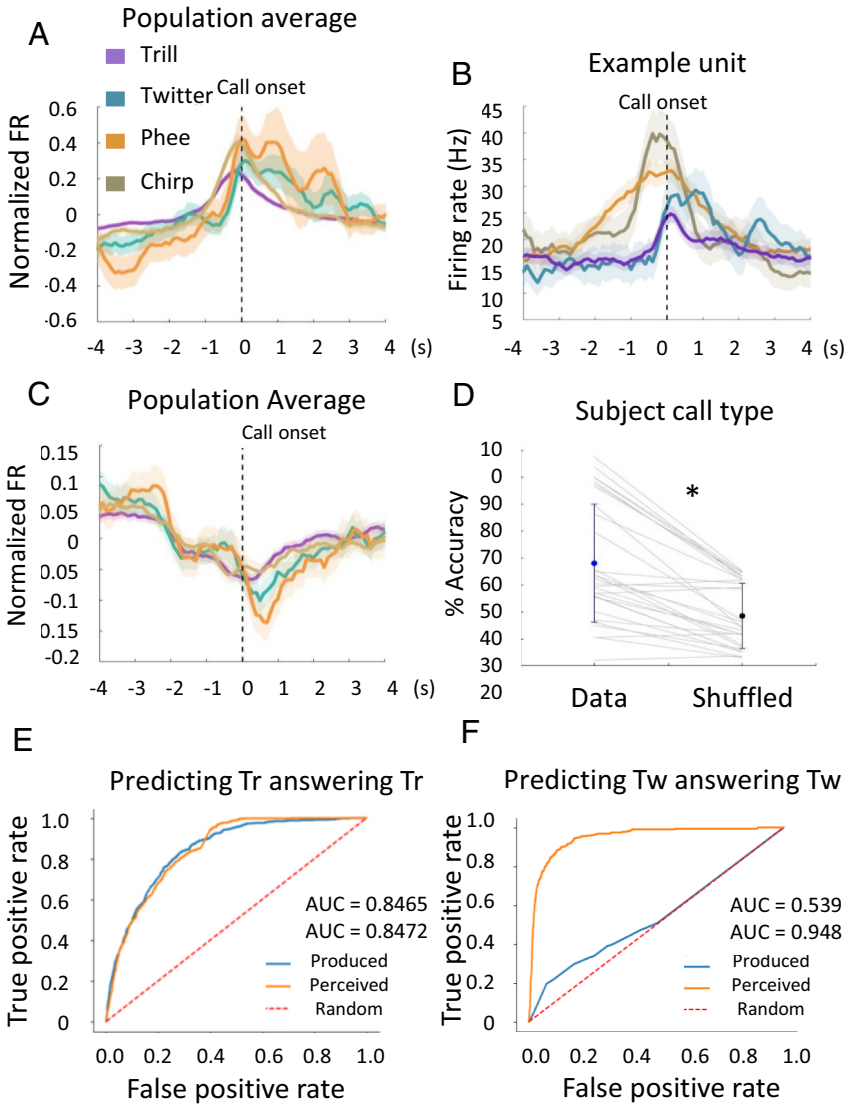


Fig. 4. ACC neurons encode active turn taking with the partner and the colony. (*A*) Normalized average increase from neurons with significantly upregulated activity during call production. (*B*) Example neuron responding to the production of several call types. (*C*) Normalized average decrease from neurons with significantly downregulated activity during call production. (*D*) Percentage accuracy of the neuronal multiclass decoder for produced call type for real data and data with shuffled labels. (*E* and *F*) Receiver Operator Characteristic (ROC) traces for our neural decoder trained on predictive time bins representing whether trills answered trills with the subject's partner in panel *E* and the answering of twitter calls with the colony in panel *F*. Blue solid lines represent calls initially produced by the subject while orange solid lines represent calls initially perceived by the subject. Random chance is represented by the red dashed line as a guide. The area under the curve (AUC) quantifies the success of the neural decoder as our outcome measure. $^*P < 0.01$, error bars indicate SEM.

vocal reply to the partner requires more attention than producing a call independently of a vocal interaction and examined if single neurons were sensitive to these contextual differences by comparing their activity in response to conversational *versus* independent calls. Analyses revealed no significant differences at the population level and only 11/183 neurons were modulated by context (*SI Appendix*, Fig. S11). Second, because marmosets typically match call types during vocal turn taking, we hypothesized that following production of a trill, attention is heightened in anticipation of a conspecific producing a trill in response. Thus, neurons sensitive to perceived trills should exhibit a stronger response than during other call types production (i.e., twitters), whereas the opposite would be true following production of a twitter. We analyzed neural activity when calls perceived were matching recently produced calls (suggesting attention, eg, partner trills within 5 seconds of a subject trill) *versus* perceived calls that did not match the recently produced call type (eg, partner trills within 5 seconds of a subject twitter) and observed no statistical differences (*SI Appendix*, Fig. S12). Although these results may suggest that attentional effects are not evident in ACC at the single unit level in this context (though see Fig. 4 *E* and *F* for population level), they also highlight the experimental challenges in identifying clear attentional effects from behavior in dynamic naturalistic contexts. As such, the absence of statistical evidence for attentional modulation in these neural analyses may not indicate that these processes are absent, but simply that these analyses are insufficient to identify attentional processes in this context.

These results indicate that ACC area 24 encodes not only the calls themselves, but information about social context. Because many of these turn-taking exchanges occurred during noisy acoustic periods, particularly for twitter calls, this also suggests that levels of background noise do not affect ACC activity during a coordinated social interaction, further emphasizing the role of this neural substrate to resolve the CPP.

Downloaded from https://www.pnas.org by UC San Diego on November 24, 2026 from IP address 64.98.200.126.

Discussion

Here, we investigated the role of ACC area 24 in resolving the CPP as freely-moving marmosets engaged in natural vocal communication in a noisy environment. We found that single neurons selectively responded to partner vocalizations or colony calls, effectively parsing different social categories of conspecific speakers in the acoustic scene (Fig. 1). Critically, neurons selective for partner calls responded invariantly when calls overlapped acoustically with colony calls, showing that neural mechanisms to distinguish meaningful voices from other interfering voices, are present in marmoset ACC area 24 (Fig. 2). These findings suggest that ACC may be a keystone neural substrate for resolving the CPP, highlighting the significance of this process for mediating social interactions in natural scenes, rather than solely a challenge for audition.

As a midline forebrain area conserved across all mammals (33) that is both anatomically connected to fronto-lateral (34) and auditory cortex (35), and modulated by attention (27–29), it may perhaps not be surprising that the ACC supports resolution of the CPP. Although not classically considered part of the language network, results here show that ACC neurons are involved in a range of communicative functions, including representing each of the call types produced or perceived in vocal exchanges (Figs. 1 and 3). Recent marmoset fMRI studies have indeed implicated ACC in vocalization processing (9, 25), and an ACC to auditory cortex pathway in gerbils has been shown to causally modulate auditory perception (36), suggesting ACC strongly interacts with the broader communication network. This indicates that ACC both encodes socio-acoustic information but also modulates it at the source, in sensory cortices. Moreover, if the CPP is viewed not as a problem of literal language comprehension but as one of identifying and tracking communicative partners in noisy social environments, resolving it requires listeners to recognize each speaker's voice. Afferent connections from hippocampus, which robustly encodes identity in social signals (32), to area 24 (35), could provide this necessary input. Taken together, these findings and the underlying anatomical connectivity suggest that ACC area 24 plays a more central role in resolving the CPP than previously thought, potentially emphasizing that the CPP is not solely a challenge for audition but one of sociality.

An alternative possibility is that the CPP is resolved in the auditory cortex, and that this processed information stream is sent to the ACC. It should be noted that auditory cortex itself also encodes both identity and noise (37), and that attended stimuli can be decoded from its activity (2, 38). However, these studies have used population signals such as LFP to decode sounds from neural activity, while here we show key mechanisms for resolving the CPP at the single cell level, such neurons have yet to be reported in studies of auditory cortex. Most critically, a recent study (36) demonstrated that without the projections from ACC to auditory cortex, auditory cortex fails to identify auditory objects, particularly in complex acoustic scenes. Overall, this suggests that the auditory cortex to ACC bidirectional network is likely integral to resolve the CPP. Combined with these previous findings, our results suggest that ACC conveys a top–down signal to the auditory cortex while simultaneously processing and abstracting auditory information from auditory sensory areas. Further research with concomitant recordings from both regions are required to validate this hypothesis.

Marmosets engage in vocal turn-taking between two or more individuals (12, 14, 39, 40). In the noisy colony setting, we observed differences in the relative timing of the call types produced in these social exchanges in the presence of background noise. Whereas mid-distance twitter calls were produced in turn-taking bouts with conspecifics in the colony, trill calls—a low amplitude call used when in close visual contact—were mostly produced during periods of silence in the colony. This shows that marmosets monitor the acoustic landscape to optimize signaling efficacy during communicative exchanges (41–44). Moreover, ACC appears to play a role in this process, as the social context in which these calls were produced and perceived could be reliably decoded from neural activity alone (Fig. 4). In parallel, fMRI findings in humans point out that ACC is activated by listening to conversations rather than monologs (45), and more generally that ACC key role in theory of mind is integral to natural language processing (46–49), and language production, as ACC lesions in humans produce akinetic mutism (50). Finally, ACC is involved in social monitoring (51), making it a logical node to be recruited for processing the CPP. Because most studies in this line of research were performed in the visual modality (10), we suggest that ACC integrates multisensory information to guide social behavior.

Growing evidence suggests broader roles for ACC in social behaviors (10, 48, 49), functions that may be particularly pertinent for resolving the CPP during natural communication. One influential framework, the Foraging Value Theory (FVT) (52), states that in a foraging for food context, the dACC is monitoring the value of the exploited option versus moving to explore other potential options. While originally developed in the context of food foraging, this framework may extend to social domains. From this perspective, the marmosets in the present study can be viewed as foraging within a social environment, dynamically adjusting their vocal interactions to maximize social reward. In humans, answering and being answered can be intrinsically rewarding (53). If similar reward contingencies apply to marmosets, area 24 may be optimizing the behavior to maximize social reward (and even perhaps in competition with other rewards/motivational factors). Consistent with this view, the tendency of marmosets to preferentially engage their partner when the colony is relatively quiet may reflect a reward-seeking strategy that complements, rather than replaces, purely acoustic optimization mechanisms. Conversely, periods of heightened colony activity may increase the relative value of interacting with other individuals, favoring broader social engagement. Because our neuronal recordings show an encoding of the social context (partner *versus* colony and answers *versus* nonanswers), area 24 function in this ecological context could be consistent with a social version of the FVT, although it should be noted that other ecological contexts might reveal additional functions of the area 24. More broadly, these findings illustrate how recording neural activity during spontaneous, naturalistic behavior can reveal functions of brain areas that remain obscured under constrained laboratory conditions. This is particularly true for ACC, which functions are debated (51).

By recording neural activity during natural vocal exchanges of freely-moving primates in rich acoustic environments, we found that the ACC area 24—a region long overlooked in this domain—implements core computations necessary for resolving the CPP. Certainly, ACC is not exclusively responsible for resolving the CPP, our results here identify it as a central, integrative hub that coordinates the perceptual, contextual, and action-related processes necessary for effective communication in noisy real-world environments. Moreover, these results demonstrate that CPP resolution extends beyond sound parsing to organizing socially meaningful interactions, requiring the integration of perception, behavioral response, and social context. This work showcases the transformative potential of naturalistic neuroscience and points to a broader reevaluation of how and where the brain solves real-world cognitive challenges (13).

Downloaded from https://www.pnas.org by UC San Diego on November 24, 2026 from IP address 64.98.200.126.

Material and Methods

Animals. Four adult marmosets (*Callithrix jacchus*) living in bonded pairs (1 female and 1 male per cage) for at least 3 mo were used for this experiment. All pairs were housed in a colony room with other marmoset cages including two adjacent cages with visual separators (between 41 and 47 marmosets in total). Three of the marmosets, monkey S (female), G (female), and L (male) were implanted with electrodes while monkey K (male) was only used for behavior. Animals had unrestricted access to food and water. All experiments were approved by the UCSD Institutional Animal Care and Use Committee and were performed in the Cortical Systems and Behavior Laboratory at University of California San Diego (UCSD).

Behavioral Paradigm. Marmosets were first habituated to be handled and taken out in transport boxes by the experimenters for a month. Once habituated to handling (measured by assessing the comfort of animals while being handled and taking treats from the experimenters' hands), we gradually habituated them over 1 or 2 mo to wear the leather harness containing the microphone until they spent less than 20% of the time chewing on it. Experimental sessions lasted between 1 h to 1.5 h. We recorded only one pair of monkeys per session. During a session, the pair was taken out of their home cage in transport boxes and brought to an adjacent experimental room for preparation. We then equipped both the subject and the partner monkey with the custom-made leather harness containing a microphone (micro voice recorder, Spycentre Security®). Additionally, we connected a neurologger to the subject's head-cap. Both monkeys were brought back to the colony room. The partner was placed back in its home-cage and the subject was kept in a transparent transport box placed at the entrance of the home-cage (Fig. 1*A* and *SI Appendix*, Fig. S2*A*) giving the pair visual and auditory access to each other. To record colony calls, we used an omnidirectional microphone (H3-VR, Zoom®) placed in the middle of the room. This paradigm allowed us to record acoustic interactions between the subject and its partner, and between the subject and the colony. In marmosets, which are capable of turn-taking, these interactions involve the coordinated exchange of vocal signals between individuals, characterized by social coordination and minimal acoustic overlap, allowing speakers to alternate the timing of their calls within a shared acoustic space.

Neurophysiological Procedures. To target the desired brain regions for electrophysiological recording using external probes, the brain anatomy of the subjects was imaged using a 3 T MRI (Siemens MAGNETOM Prisma) scanner and a human knee coil (Siemens/QED Transmit/Receive 15 channels). T2 and T1 weighted images of the whole marmoset brain were obtained in an anesthetized animal placed in a custom nonmagnetic 3D printed stereotax with earbars and bite plate. We adapted the coordinates from the Paxinos atlas (54) of the marmoset brain based on individual brain size. We targeted the left hemisphere area 24 (Site 1: AP +12 mm, ML +1 mm, Site 2: AP +13.5 mm, ML +1 mm). On the day of surgery, monkeys were intubated and maintained under isoflurane anesthesia, then placed in a surgical stereotaxic frame (Kopf). We performed a skin incision and a craniotomy (1.5 mm diameter), then inserted a ground electrode between the dura and the skull before covering the skull with metabond dental cement (C&B Metabond®). We employed 2 different strategies. For monkey S and G, we used micro wire brush arrays (Microprobes for Life Science, Maryland, USA, Monkey S:1x64 channels, Monkey G: 2 × 32 channels with 2 mm space in between) with microdrives [custom made from (55), for Monkey G, an additional hole was pierced and the drive adapted to hold 2 micro brush arrays]. The electrode guide tube(s) (Microlumen, High performance medical tubing, 310-I.5 PTFE), containing a 24G needle, attached to the microdrive was/were inserted straight into the brain at a depth of 1.5 mm. The needle(s) was/were retracted and the electrode(s) array(s) were lowered into the guide tube(s) until they protruded out of the guide tube by 1 mm. Kwik Sil (World Precision Instrument) was used to fill the craniotomy and seal the gap at the top of the guide tube(s). Then the Microdrive and the omnetics connector were cemented to the metabond using dental acrylic, after soldering the ground electrode to the array ground wire. Finally, the skin was glued to the metabond using Vetbond (3 M). For monkey L, we employed Spikegadget wireless neuropixel system. The procedure was similar, except that a durotomy was performed and the 2 probes (Npx 1.0 and Npx NHP 1.0 10 mm) inserted 5 mm deep from the cortical surface before we covered them with kiwk-Sil and fixed them directly to the skull with metabond. Probes were then connected to the Spikegadget head cap that was built around the electrodes and cemented to the metabond. After the surgery, monkeys received analgesic treatment for 3 d. Recording sessions started 2 wk after the surgery for monkey S and G and were performed twice per week, and started 4 d after surgery and happened every day for monkey L (monkey S: 21 sessions, monkey G: 19 sessions, Monkey L: 6 sessions). Brush arrays were lowered thanks to the micro drive every 3 or 4 sessions to record fresh neurons. Implantation sites were verified by structural MRI for monkey G and S (*SI Appendix*, Fig. S1*B*), but this could not be done for monkey L.

To wirelessly record neurons, we used a 64 channels Neurologger (Deuteron Technologies) for monkey S and G with a sample rate of 32 KHz, and Neuropixels Datalogger Headstage (SpikeGadgets) with a sample rate of 30 KHz for monkey L.

Data Preprocessing.

Vocalizations. To automatically label subjects' and partners' vocalizations, we retrained the algorithm described in ref. 20 on the data recorded from our 4 monkeys with the wearable microphones. This artificial neural network (56) identifies the call type and determines the source (subject or partner) by comparing the amplitude. If the amplitude is equal in both microphones, the call is considered to be coming from the colony. Following this step, all audio files were manually corrected to remove eventual mislabeling by the algorithm and determine precisely (<20 ms) each call onset and offset. We considered only 4 call types: trill, twitter, phee, and chirp. Other calls (tsik, ek, chatter, trillPhee, whistle) had too few occurrences to be analyzed. Additionally, to investigate the cocktail party effect, Twitter calls coming from the partner were labeled as "simultaneous" or "sequential" depending on if a call from the colony was overlapping or not with it.

For calls coming from the colony (recorded with the omnidirectional microphone), a first pass was done with ACDC neural network (https://github.com/mineraldragon/ACDC_2022) and then corrected manually. We only labeled Twitter and Phee calls from the colony, as other call types such as trill and chirp were not loud enough to be recorded consistently, or had too few occurrences of them (e.g., tsik, chatter, etc.). Note that for technical reasons, colony calls were not always recorded (34 out of 46 sessions with colony recordings).

We removed all calls that were happening less than 2 s after another call from the same source, to ensure that neuronal responses analyzed were not influenced by a recent prior call. This was the case for all analyses except for PSTH of phee calls perceived from the partner (Fig. 1*F*), due to low number of calls. Thus, Fig. 1*F* phee call data may contain residual activity from successive calls.

Finally, we removed calls that occurred when the neural signal was poor due to artifacts, defined as calls for which 20% of the neural signal (from 1 s before to 1 s after the call onset) was labeled as artifact (*Neural Data*).

Neural data. All neural data were band pass filtered (300 to 7,000 Hz). Signal from brush arrays (but not Neuropixel) contained artifacts, that were removed by setting signal value to 0 if the SD of one channel was above a threshold (arbitrarily set to 100 after visual inspection of signal). If more than 20% of a session was removed, the session was not included in our analyses. We then used Kilosort 2 (57) to perform automatically the spike sorting for both micro brush arrays and Neuropixels data. Each identified unit was manually checked and validated and labeled as single or multiunit or rejected based on the waveform shape and auto-correlogram using Phy (open-source Python library for spike visualization and curation). Only units with a minimum firing rate of 0.5 Hz were considered. Across the 3 monkeys (Monkey S: 231 units, Monkey G: 574 units, Monkey L: 794 units) a total of 1,599 units were identified, of which 932 were well isolated single units (using the PCA). However, because we did not see differences between single and multiunits in later analyses, we decided to pool them together and to refer to them (single and multiunits) as neurons throughout this article. All PSTH in this article are showing results from well isolated single units. Note that some units may have been present on several sessions, however, because we regularly lowered our brush arrays and based on previous work, the majority of units (>50%) should have been recorded only once (55, 58, 59). Furthermore, we found no differences between the response profile of individual monkeys (*SI Appendix*, Fig. S3*A*), and the proportions of responses to each call type production and perception were similar for sessions recorded with brush arrays at the same depth (22.7%), after a push (26.1%), or with neuropixel (33.5%), indicating that even if some neurons were recorded twice, this did not affect the overall response profile of our neuronal populations.

Downloaded from https://www.pnas.org by UC San Diego on November 24, 2026 from IP address 64.98.200.126.

Synchronization. All data streams (3 microphones and neurologger) were synchronized using 250 ms audio tone pulses played via a buzzer generated by an ESP32 microcontroller programmed using Arduino IDE.

All .wave files were manually aligned with Audacity® (v 3.4). Two tones (one at the start of the session and another at the end) were used to bookend the whole session which allowed us to identify and correct the temporal drift of each microphone in Audacity using the "change speed" functionality. Audacity was able to correct the temporal drift in each microphone to less than 20 ms. Following this, a MATLAB script was used to randomly remove data points until the temporal time drift was reduced to <1 ms. This did not influence the labeling of calls by the artificial neural network.

A 100 mv TTL pulse (100 ms) was sent to the Neurologger using the same ESP32 microcontroller neuron at the beginning of the session. The time difference between the Neurologger TTL and first acoustic pulse was used to align the timing of vocalizations onset and offset to the neurologger time. Using this approach, we were able to synchronize different systems with varying sampling frequencies to a common master clock.

Statistical Analyses.

PSTH responses. Firing rate was computed over 100 ms bins. To test neurons' response to produced or perceived calls we first employed a peak detection approach. For each call type, if there were at least 5 calls, we calculated the mean firing rate between −5 s to −2 s before call onset for produced calls and between −5 s to 0 s before call onset for perceived calls as a baseline. We then searched for the peak response between −2 s and +1 s around call onset for produced calls and between 0 s and 3 s after call onset for calls perceived. To ensure a minimum peak width, we then tested if the bins around the peak bin reached at least 50% of the peak value [threshold = baseline + −0.5 * (peak-baseline)] to have at least 3 consecutive bins (300 ms) beyond this threshold. Finally, we tested if the peak values were significantly higher than the baseline values using the nonparametric paired Wilcoxon signed rank test. We used this approach to allow for time flexibility because we observed that neurons responded at various time windows (Fig. 4 and *SI Appendix*, Fig. S6). Note that both maximum and minimum peak responses were computed to look for excited and inhibited neurons. Each neuron was tested for each produced and perceived call types.

Finally, the PSTH traces reported in *SI Appendix*, Fig. S6 were first computed at a bin width of 0.3 s, then up-sampled to a bin width of 0.1 s using linear interpolation that assigned knots to the mean time of any given bin, and finally smoothed with a sliding window with a duration of 2 s using a Savitzky–Golay filtration. Uncertainty of these demonstrational PSTH traces indicates 95% confidence estimated via bootstrap and is shown by the relatively horizontal, yet curved shaded regions in *SI Appendix*, Fig. S6.

Decoding. To decode call type or caller category, we employed a multiclass decoder (error-correcting output codes model, MATLAB function *fitcecoc*) with a one-versus-all coding design. For each session, the input was the average firing rate of each neuron between −2 s and +1 s around call onset for produced calls and between 0 s and 3 s after call onset for calls perceived. The number of calls of each type (trill, twitter, phee, or chirp) or caller (partner or colony) was equalized by subsampling categories with more calls down to the category that had the least. Only calls that had at least 10 occurrences were considered, leading to different number of categories and therefore chance levels for each session (25%, 33%, or 50%). The decoder was trained on 80% of the data and tested on the remaining 20%. To control for random effects, we also decoded on neural data with shuffled labels. This procedure was repeated 100 times to obtain a representative average decoding value for the data and the shuffle. Finally, we used Wilcoxon sign rank test to compare the average decoding accuracies obtained on the real and the shuffled data.

Time bins. To quantify the neural representations that identified whether a trill with the partner was answered or whether a twitter with the colony was answered, we used a predictive time bin analysis similar to previous work in the lab (32, 60). In this study, we applied the same four-fold stratified cross-validation both to detect predictive time bins and to decode their apparent firing rates using an ensemble of gradient-boosted decision trees (61). Recording sessions were selected for these analyses if they had at least ten repeatedly spiking neurons recorded, at least four recorded calls that were answered, and at least four recorded calls that were not answered. A trill call was considered an answer if it happened within 5 s of a trill call from the other monkey, and a twitter call was considered an answer if it happened within 5 s of a twitter call from other colony monkeys. These timings were based on previous literature (20, 62). As a control, the same analysis was repeated with pseudorandomly shuffled labels, and decoding results are reported in the main text and in Fig. 4. All time bins considered only the times from five seconds before to 5 s after call onset and were constrained to be no briefer than four hundred milliseconds in duration, resulting in a median activation time of t = -0.17 (IQR: −2.34 through +2.14) seconds (N_{bins} = 4,326 predictive time bins) relative to call onset at t = 0. Predictive time bins were also constrained to be nonoverlapping for any given neuron such that no action potential was considered twice for a given type of neural representation. Reported tree-based decoder accuracies were computed using a threshold parameter value that maximized apparent accuracy. Area under the ROC curves provide an alternative outcome measure that is parameterless and is reported in the Fig. 4 *E* and *F* and *SI Appendix*, Fig. S12 *C* and *D*.

We found 2,139 significant predictive time bins across 948 neurons that were significantly predictive if a trill call was part of a turn-taking bout in terms of median firing rate (two-sided Mann–Whitney's *U* test, $P < 0.05$, *SI Appendix*, Fig. S12 *A* and *B*). Only 1,098 neurons were considered in these trill call analyses by the minimal session selection criteria, which was used equally in all predictive time bin analyses. We found 1,545 predictive time bins across 417 neurons that were significantly predictive if a twitter call was part of a turn-taking bout (two-sided Mann–Whitney's *U* test, $P < 0.05$). Only 682 neurons were considered in these twitter call analyses by the same minimal session selection criteria.

Data, Materials, and Software Availability. This data can be shared upon request but are not made publicly available on a repository because they legally belong to both the CNRS (France) and UCSD (USA), which have different policies on data sharing. All data and code are available on request to the corresponding author (Arthur.lefevre [@] isc.cnrs.fr) or by contacting the following email addresses: comiller [@] ucsd.edu or isc [@] isc.cnrs.fr.

ACKNOWLEDGMENTS. This work is supported by Grants to A.L. (Marie Sklodowska-Curie fellowship 101018877 and European Research Council Starting Grant 101116110) and to C.T.M. (National Institute of Health R01 DC 012087).

Author affiliations: [a]Cortical Systems and Behavior Laboratory, Department of Psychology, University of California San Diego, La Jolla, CA 92093; [b]Institute of Cognitive Sciences Marc Jeannerod, CNRS, University Lyon 1, Bron 69500, France; and [c]Neuroscience Graduate Program, University of California San Diego, La Jolla, CA 92093

1. J. H. McDermott, The cocktail party problem. *Curr. Biol.* **19**, R1024–R1027 (2009).
2. N. Mesgarani, E. F. Chang, Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* **485**, 233–236 (2012).
3. M. A. Bee, C. Micheyl, The cocktail party problem: What is it? How can it be solved? And why should animal behaviorists study it?. *J. Comp. Psychol.* **1983**, 235–251 (2008).
4. J. O'Sullivan *et al.*, Hierarchical encoding of attended auditory objects in multi-talker speech perception. *Neuron* **104**, 1195–1209.e3 (2019).
5. I. Nelken, J. Bizley, S. A. Shamma, X. Wang, Auditory cortical processing in real-world listening: The auditory system going real. *J. Neurosci.* **34**, 15135–15138 (2014).
6. N. Joshi *et al.*, Temporal coherence shapes cortical responses to speech mixtures in a ferret cocktail party. *Commun. Biol.* **7**, 1392 (2024).
7. J. T. Klein, S. V. Shepherd, M. L. Platt, Social attention and the brain. *Curr. Biol.* **19**, R958–R962 (2009).
8. U. Jürgens, Neural pathways underlying vocal control. *Neurosci. Biobehav. Rev.* **26**, 235–258 (2002).
9. A. Jafari *et al.*, A vocalization-processing network in marmosets. *Cell Rep.* **42**, 112526 (2023).
10. P. T. Putnam, S. W. C. Chang, Social processing by the primate medial frontal cortex. *Int. Rev. Neurobiol.* **158**, 213–248 (2021).
11. A. Nieder, R. Mooney, The neurobiology of innate, volitional and learned vocalizations in mammals and birds. *Philos. Trans. R. Soc. B Biol. Sci.* **375**, 20190054 (2020).
12. J. M. Burkart *et al.*, A convergent interaction engine: Vocal communication among marmoset monkeys. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **377**, 20210098 (2022).
13. C. T. Miller *et al.*, Natural behavior is the language of the brain. *Curr. Biol.* **32**, R482–R493 (2022).
14. B. M. Bezerra, A. Souto, Structure and usage of the vocal repertoire of Callithrix jacchus. *Int. J. Primatol.* **29**, 671–701 (2008).
15. L. J. Digby, C. E. Barreto, Social organization in a wild population of *Callithrix jacchus*. I. Group composition and dynamics. *Folia Primatol. Int. J. Primatol.* **61**, 123–134 (1993).
16. C. Lazaro-Perea, Intergroup interactions in wild common marmosets, *Callithrix jacchus*: Territorial defence and assessment of neighbours. *Anim. Behav.* **62**, 11–21 (2001).

Downloaded from https://www.pnas.org by UC San Diego on November 24, 2026 from IP address 64.98.200.126.

17. C. T. Miller *et al.*, Marmosets: A neuroscientific model of human social behavior. *Neuron* **90**, 219–233 (2016).
18. J. M. Burkart, C. P. van Schaik, Marmoset prosociality is intentional. *Anim. Cogn.* **23**, 581–594 (2020).
19. A. Vitale, M. Zanzoni, A. Queyras, F. Chiarotti, Degree of social contact affects the emission of food calls in the common marmoset (Callithrix jacchus). *Am. J. Primatol.* **59**, 21–28 (2003).
20. R. Landman *et al.*, Close-range vocal interaction in the common marmoset (Callithrix jacchus). *PloS ONE* **15**, e0227392 (2020).
21. B. W. Robinson, Vocalization evoked from forebrain in Macaca mulatta. *Physiol. Behav.* **2**, 345–354 (1967).
22. F. Sperli, L. Spinelli, C. Pollo, M. Seeck, Contralateral smile and laughter, but no mirth, induced by electrical stimulation of the cingulate cortex. *Epilepsia* **47**, 440–443 (2006).
23. N. Gavrilov, S. R. Hage, A. Nieder, Functional specialization of the primate frontal lobe during cognitive control of vocalizations. *Cell Rep.* **21**, 2393–2406 (2017).
24. R. A. West, C. R. Larson, Neurons of the anterior mesial cortex related to faciovocal activity in the awake monkey. *J. Neurophysiol.* **74**, 1856–1869 (1995).
25. A. Dureux, A. Zanini, R. Trapeau, P. Belin, S. Everling, Functional organization of voice patches in marmosets and cross-species comparisons with macaques and humans. *Curr. Biol.* **35**, 3869–3882.e4 (2025).
26. A. S. Bregman, Auditory scene analysis: Hearing in complex environments (MIT Press, 1993), pp. 1994–790.
27. S. E. Petersen, M. I. Posner, The attention system of the human brain: 20 years after. *Annu. Rev. Neurosci.* **35**, 73–89 (2012).
28. K. N. Schneider, X. A. Sciarillo, J. L. Nudelman, J. F. Cheer, M. R. Roesch, Anterior cingulate cortex signals attention in a social paradigm that manipulates reward and shock. *Curr. Biol.* **30**, 3724–3735.e2 (2020).
29. R. H. B. Benedict *et al.*, Covert auditory attention generates activation in the rostral/dorsal anterior cingulate cortex. *J. Cogn. Neurosci.* **14**, 637–645 (2002).
30. H.-C. Chen, G. Kaplan, L. J. Rogers, Contact calls of common marmosets (*Callithrix jacchus*): Influence of age of caller on antiphonal calling and other vocal responses. *Am. J. Primatol.* **71**, 165–170 (2009).
31. J. Li, M. C. Aoi, C. T. Miller, Representing the dynamics of natural marmoset vocal behaviors in frontal cortex. *Neuron* **112**, 3542–3550.e3 (2024).
32. T. J. Tyree, M. Metke, C. T. Miller, Cross-modal representation of identity in the primate hippocampus. *Science* **382**, 417–423 (2023).
33. A. Burgos-Robles, K. M. Gothard, M. H. Monfils, A. Morozov, A. Vicentic, Conserved features of anterior cingulate networks support observational learning across species. *Neurosci. Biobehav. Rev.* **107**, 215–228 (2019).
34. M. Ducret *et al.*, Medial to lateral frontal functional connectivity mapping reveals the organization of cingulate cortex. *Cereb. Cortex N.Y. 1991* **34**, bhae322 (2024).
35. B. A. Vogt, D. N. Pandya, Cingulate cortex of the rhesus monkey: II. Cortical afferents. *J. Comp. Neurol.* **262**, 271–289 (1987).
36. K. L. Anbuhl, M. Diez Castro, N. A. Lee, V. S. Lee, D. H. Sanes, The cingulate cortex facilitates auditory perception under challenging listening conditions. *Proc. Natl. Acad. Sci. U.S.A.* **122**, e2412453122 (2025).
37. S. M. Town, K. C. Wood, J. K. Bizley, Sound identity is represented robustly in auditory cortex during perceptual constancy. *Nat. Commun.* **9**, 4786 (2018).
38. J. Z. Simon, The encoding of auditory objects in auditory cortex: Insights from magnetoencephalography. *Int. J. Psychophysiol.* **95**, 184–190 (2015).
39. A. B. Bosshard *et al.*, Beyond bigrams: Call sequencing in the common marmoset (Callithrix jacchus) vocal system. *R. Soc. Open Sci.* **11**, 240218 (2024).
40. G. Oren *et al.*, Vocal labeling of others by nonhuman primates. *Science* **385**, 996–1003 (2024).
41. S. J. Eliades, X. Wang, Neural correlates of the lombard effect in primate auditory cortex. *J. Neurosci.* **32**, 10737–10748 (2012).
42. J. Tsunada, S. J. Eliades, Frontal-auditory cortical interactions and sensory prediction during vocal production in marmoset monkeys. *Curr. Biol.* **35**, 2307–2307.e3 (2025).
43. J. Löschner, T. Pomberger, S. R. Hage, Marmoset monkeys use different avoidance strategies to cope with ambient noise during vocal behavior. *iScience* **26**, 106219 (2023).
44. J. Löschner, S. R. Hage, Sound amongst the din: Primate strategies against noise. *Trends Cogn. Sci.* **29**, 111–113 (2025).
45. H. A. Olson, E. M. Chen, K. O. Lydic, R. R. Saxe, Left-hemisphere cortical language regions respond equally to observed dialogue and monologue. *Neurobiol. Lang. Camb. Mass* **4**, 575–610 (2023).
46. E. Fedorenko, A. A. Ivanova, T. I. Regev, The language network as a natural kind within the broader landscape of the human brain. *Nat. Rev. Neurosci.* **25**, 289–312 (2024).
47. E. C. Ferstl, D. Y. von Cramon, What does the frontomedian cortex contribute to language processing: Coherence or theory of mind? *Neuroimage* **17**, 1599–1612 (2002).
48. D. M. Amodio, C. D. Frith, Meeting of minds: The medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* **7**, 268–277 (2006).
49. M. K. Wittmann, P. L. Lockwood, M. F. S. Rushworth, Neural mechanisms of social cognition in primates. *Annu. Rev. Neurosci.* **41**, 99–118 (2018).
50. R. W. Barris, H. R. Schuman, Bilateral anterior cingulate gyrus lesions; Syndrome of the anterior cingulate gyri. *Neurology* **3**, 44–52 (1953).
51. N. Clairis, A. Lopez-Persem, Debates on the dorsomedial prefrontal/dorsal anterior cingulate cortex: Insights for future research. *Brain* **146**, 4826–4844 (2023).
52. N. Kolling, T. E. Behrens, R. B. Mars, M. F. Rushworth, Neural mechanisms of foraging. *Science* **336**, 95 (2012).
53. D. A. Abrams, P. K. Mistry, A. E. Baker, A. Padmanabhan, V. Menon, A neurodevelopmental shift in reward circuitry from mother's to nonfamilial voices in adolescence. *J. Neurosci.* **42**, 4164–4173 (2022).
54. G. Paxinos, C. Watson, M. Petrides, M. Rosa, H. Tokuno, *The Marmoset Brain in Stereotaxic Coordinates* (Elsevier Academic Press, 2012).
55. D. B. T. McMahon, I. V. Bondar, O. A. T. Afuwape, D. C. Ide, D. A. Leopold, One month in the life of a neuron: Longitudinal single-unit electrophysiology in the monkey visual system. *J. Neurophysiol.* **112**, 1748–1762 (2014).
56. T. Oikarinen *et al.*, Deep convolutional network for animal sound classification and source attribution using dual audio recordings. *J. Acoust. Soc. Am.* **145**, 654 (2019).
57. M. Pachitariu, N. Steinmetz, S. Kadir, M. Carandini, K. D. Harris, Kilosort: Realtime spike-sorting for extracellular electrophysiology with hundreds of channels. bioRxiv [Preprint] (2016). https://www.biorxiv.org/content/10.1101/061481v1 [Accessed 12 August 2025].
58. A. X. Yuan *et al.*, Multi-day neuron tracking in high-density electrophysiology recordings using earth mover's distance. *eLife* **12**, RP92495 (2024).
59. E. H. van Beest *et al.*, Tracking neurons across days with high-density probes. *Nat. Methods* **22**, 778–787 (2025).
60. T. J. Tyree, Applications of mathematical physics to quantitative biology (Dissertations Abstracts International, 2023), vol. 85, p. 167.
61. T. Chen, C. Guestrin, "XGBoost: A scalable tree boosting system" in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data MiningPages* (Association for Computing Machinery, New York, NY, 2016), pp. 785–794.
62. V. Jovanovic, A. R. Fishbein, L. de la Mothe, K.-F. Lee, C. T. Miller, Behavioral context affects social signal representations within single primate prefrontal cortex neurons. *Neuron* **110**, 1318–1326.e4 (2022).

Downloaded from https://www.pnas.org by UC San Diego on November 24, 2026 from IP address 64.98.200.126.