Active sensing in vocal interactions

Introduction

Most behavioral neuroscience research focuses on how inputs are perceived or how animals respond to them, but this approach considers the input as a passive process. This is problematic because animals sometimes behave in a way that purposely changes the sensory information they will receive, a process called active sensing. There are multiple examples of active sensing described in the literature (Zweifel & Hartmann, 2020 [Defining "active sensing" through an analysis of sensing energetics: homeoactive and alloactive sensing - PubMed (nih.gov)](#) ; Jones et al., 2021 [Communication with self, friends and foes in active-sensing animals - PubMed (nih.gov)](#); Luo et al., 2021 [Constraints on the deformation of the vibrissa within the follicle | PLOS Computational Biology](#)).

==~ Human example ~==

One of the most popular example being currently studied by neuroscientists is the rat's use of their whiskers . Rats are constantly moving their highly sensitive whiskers across the environment in a sweeping motion, and depending on how each whisker resonates with the surface it is in contact with, a different signal goes to the rat's body, where it is processed to help the rat explore its surroundings. This is a sensory modality that is mostly specific to animals with long whiskers, but when you consider the fact that most animals point their head in a way that they can see something specific, which is called active vision (Bajcsy, 1988 [Active perception - Proceedings of the IEEE (berkeley.edu)](#)), or moving their limbs around an specific object, called active touch (Gibson, 1962, [Gibson (1962) Observations on active touch (zhdk.ch)](#)), we can say that all of these animals do active sensing, including humans. Even hearing, that is generally thought to be a very passive process, has a well described active sensing literature within echolocation (Beleyur, 2019 [Modeling active sensing reveals echo detection even in large groups of bats | PNAS](#)). Bats send an ultrasound signal and can hear back the reflected sound, helping them visualize their surroundings in the dark.

Active hearing is not specific for bats, however. Vocal interactions are an important way for animals to exchange information, for instance communicating the presence of a predator to a group. Emitting a vocalization not only transmits information to other individuals, but is often a way of seeking information from individuals when there is a

response, particularly when the two individuals are not within each other's visual field. We know that frogs exhibit a high selectivity to what calls elicit a response (Capranica, 1965 https://psycnet.apa.org/record/1965-14349-000). In particular, in some frogs, different call durations have different likelihoods to elicit a response (Penna, 1997 https://doi.org/10.2307/1565388). That brings up the question of whether other animals, such as marmosets, make their calls to optimize for response likelihood or something else related, such as maximizing how much environment they can extract from the environment.

There are energetic costs and time-limitations to producing vocalizations in a way that generates a response, so there should be a balance between the potential information obtained from vocalizing and the costs associated with this process. This problem has been characterized before when animals use their movement for sensing (MacIver et al, 2010 Energy-Information Trade-Offs between Movement and Sensing | PLOS Computational Biology ; Chen et al., 2020 Tuning movement for sensing in an uncertain world | eLife (elifesciences.org)). Adapting the model used for movement, we can describe a way in which vocalizations are used to simultaneously optimize information acquisition and potentially the energy expenditure. In the navigation example, the model predicts an efficient exploration of the space. Here, we expect the exploration of the vocalization space, exhibiting diversity in the vocal repertoire in a single session, even when there is a specific vocalization that appears to be optimal in that session. Such vocal diversity is present in many species, such as marmoset monkeys, songbirds, and humans (refs).

In this paper, we chose the common marmoset (*Callithrix jacchus*) as a test for this model. We know that marmoset's use of their vocalization is also characterized by a balance between the energy expenditure to vocalize and the potential information it can obtain (Varella et al., 2022). Marmosets have a very diverse vocal repertoire, including vocalizations that are most commonly observed in specific contexts, like the phee calls used to contact far away conspecifics, or trills to keep a group together (Bezerra & Souto, 2008 Structure and Usage of the Vocal Repertoire of Callithrix jacchus | SpringerLink). A vocal accommodation framework was put forward (Ruch et al., 2017 The function and mechanism of vocal accommodation in humans and other primates (wiley.com)) describing that vocalizations can be adapted to certain contexts to signal social closeness and to optimize signal transmission. If a vocalization is not optimized to these factors, a vocal exchange may not be established, and the call will not be followed by a response, suggesting that these factors are related to the optimization of the probability of vocal response. It is not well understood, however, why marmosets emit different call types in contexts in which that vocalization may not be the optimal. In

the model presented in this paper, we suggest that these apparently suboptimal vocalizations are necessary to continually probe the environment as an active sensing mechanism.

Results

In this study, we recorded vocalizations from six marmosets, grouped into three pairs. For each recording, we extracted the onset and offset of each vocalization, the amplitude of the call, the caller ID, and the type of the call. The marmoset vocal repertoire includes various types of vocalizations, with affiliative calls (such as phees, trillphees, and trills) predominating in situations where the two marmosets are visually occluded from each other (as depicted in Figure 1a) (Liao et al., 2018). In this context, marmosets have been observed to engage in turn-taking behavior, similar to that of coupled oscillators (Takahashi et al., 2013), characterized by consistent periods between vocalizations. For a successful turn-taking dynamics, the marmosets must have an expectation of some features of the other individual's vocalization, for example, how long the incoming vocalization might last.

We analyzed the correlation between calls and responses to determine the source of the information used by the marmoset to generate its expectations. We considered three possibilities: that the information is derived from the previous call, from elements that define the session, or that it is the same information every time, in which case there is no learning from the session or previous call. Based on previous literature (Takahashi et al., 2013), we defined a vocalization as a response if it was a vocalization from a different marmoset that occurred within 12 seconds of the first vocalization. This allowed us to generate a dataset of call and response durations (as shown in Figure 1b).
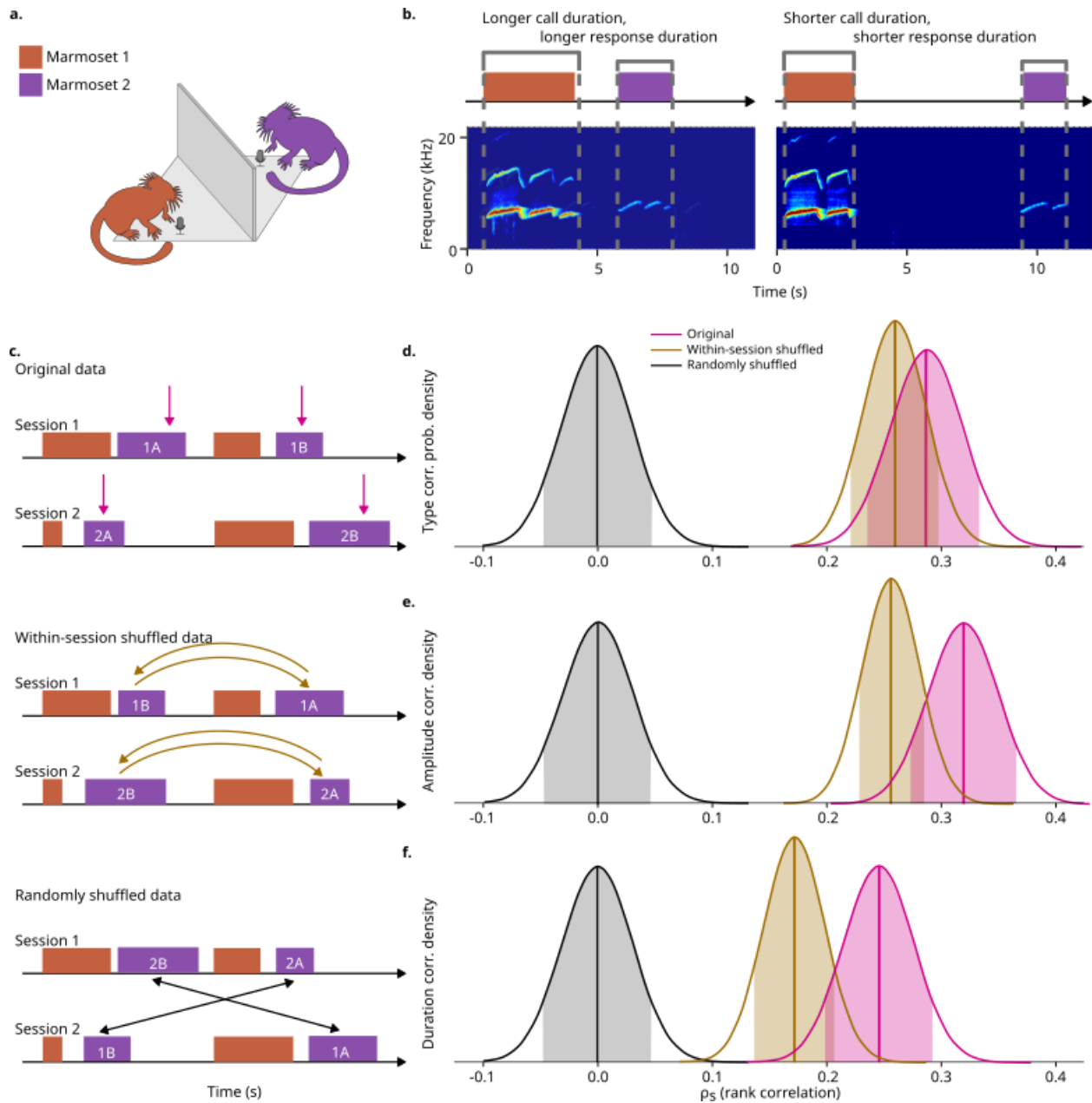
1. Response strategy depends on the session. a. Schematics of experimental set-up. **b.** Spectrogram exemplifying call and response durations. **c.** Schematics of shuffling of call and response. In the original data, the calls and responses are maintained as they were. In the "within-session shuffled data", vocal responses of the same individual are shuffled within each recording session. In the "randomly shuffled data", vocal responses are shuffled regardless of who the individual is or what session it is. (**d-f**). Kernel density estimation (probability density) of the Spearman's correlation coefficient (rank correlation) of call type, amplitude, and duration, respectively, between a call and a response. All group differences are significant ($p < 0.001$) except for within-session shuffled data and original data for call type, in **a**. Shaded areas represent a bootstrapped 95% confidence interval. This graph shows a significantly higher correlation between call and response features when compared to a random distribution of call-responses pairs, suggesting there is shared information between the response and the call. There is also information contained in the session.

We found that there's a significant amount of information about the response on the previous call, although most of that information is also present in the session as a whole (Figure 1d-f). We looked at the relationship between the call and response by analyzing three different characteristics of the vocalizations (the type of call, the call amplitude, and the call duration). We found that all of these characteristics had a strong positive relationship with the response. (Spearman ρ_s of 0.29 [0.24,0.33], 0.32 [0.27,0.36], and 0.25 [0.20, 0.29], respectively. $p < 0.001$ for all of them). To control for the possibility that all marmoset vocalizations are correlated, regardless of the session or the previous call, we randomly shuffled these pairs (Figure 1c, bottom panel), and bootstrapped the correlation between calls and random responses. For all 3 vocal features, the correlations were not significantly different than 0, and were all significantly smaller than the non-shuffled data ($p < 0.001$). We can estimate how much information is coming from the session instead of the specific pair call-response by shuffling the responses emitted by one specific marmoset within a specific session (Figure 1c, middle panel). The correlations obtained for call type, amplitude, and duration, were 0.26 [0.22, 0.29], 0.26 [0.23, 0.28], and 0.17 [0.14, 0.20], respectively. All of these correlations were significantly higher than the randomly shuffled correlations ($p < 0.001$), and significantly lower than the original data, except for the call type. The call type correlation for within-sessions shuffling was not significantly smaller than the correlations for the original data ($p = 0.093$). That means that, in this context, almost none of the information that the marmoset uses to determine the call type emitted as a response is particular to the previous call - the information is embedded in the session, such as the distance between the two marmosets, the identities of the individuals, the environment they are in, etc.

The theory of vocal accommodation claims that marmosets optimize for signal transmission and ultimately communication (environmental accommodation, Ruch et al., 2018). That provides an explanation for why the information is contained in the session - there must be a set of features that are optimally associated with a particular session, for example by maximizing the amount of communication (i.e. the response likelihood) during a vocal exchange. For example, if a vocalization is too soft, the other marmoset might not hear it. If it is too loud, it could for example signify an inappropriate level of arousal and startle the other marmoset.
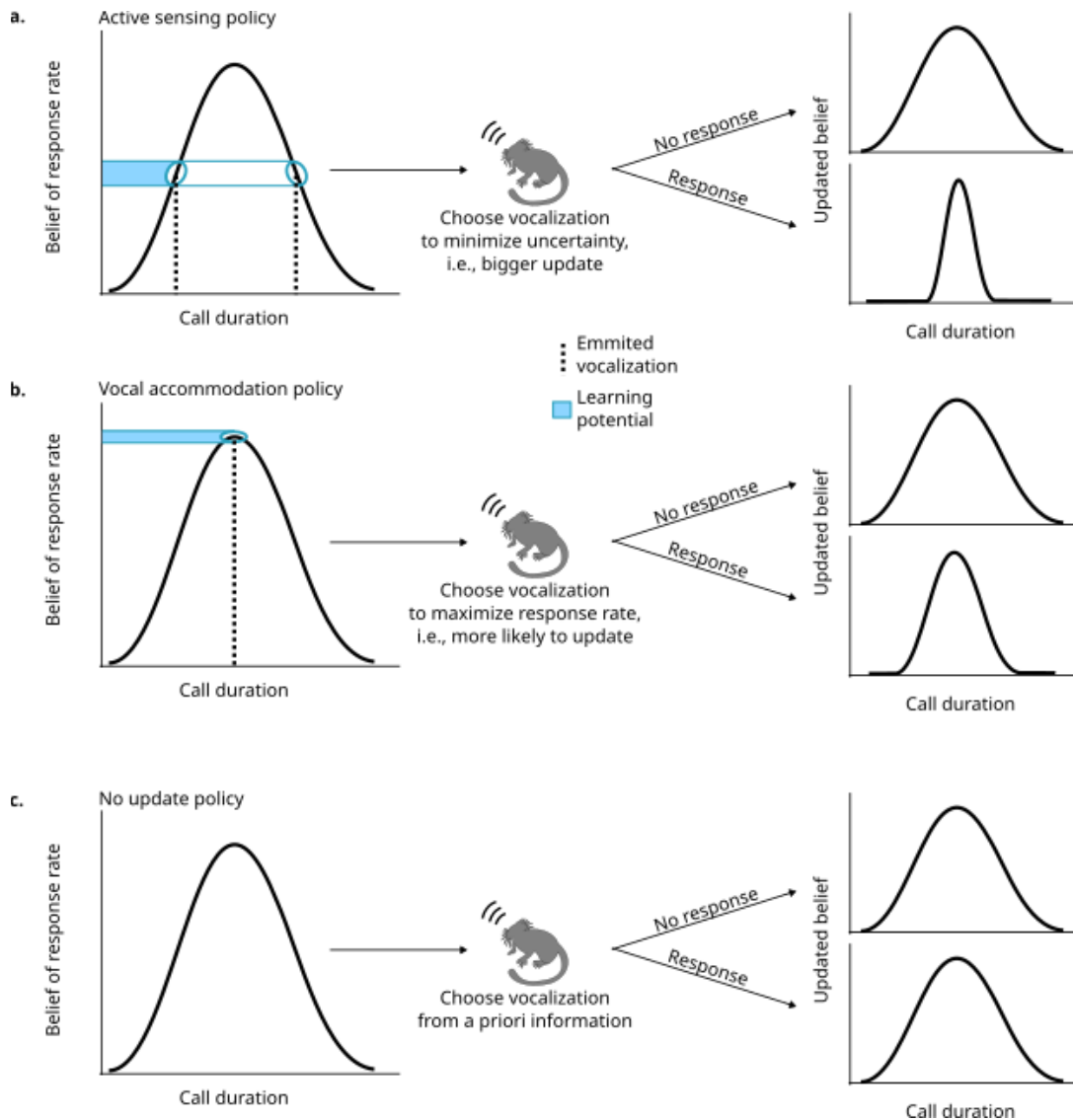
2. Social sensing models. In the model, the marmoset starts with a belief of what vocalization features lead to the highest response likelihood. Using a policy applied to this belief (i.e. the rules determining what action to take), the marmoset emits a vocalization that has specific features. Based on the information it gets from hearing or not a response, it might update the belief using the Bayes rule. **a.** In the no-update policy represented here, the belief does not change regardless of hearing a response. **b.** One possibility for a policy associated with updating the belief is a vocal accommodation policy, in which the marmoset chooses the vocalization that maximizes the response likelihood. **c.** Another possibility is the active sensing policy, using the expected information density (EID, see methods). In this case, the vocalization is chosen in a way that minimizes the uncertainty, leading to a narrower belief when it is able to make an update. In the example in the figure, the policy leads to vocalizations around the peak, rather than the peak, because they have a higher learning potential, achieving an update in the belief with higher magnitude.

Since there are so many variables that constitute a session, marmosets have multiple sensory systems - such as vision, olfaction, and audition - to extract information from the session and then use that information to adapt how they are vocalizing. In contexts where the two interacting individuals are too distant to feel, smell, or see each other, or

with objects in between, such as trees that occlude their vision, the audition becomes of particular importance to gather information about the other individual (and, by extension, how to best vocalize in a particular context). To use the audition, there must be another marmoset vocalizing, and an effective way of triggering a vocalization from another individual is by vocalizing in the first place (see Takahashi et al. 2013; Borjon et al., 2016). Figure 2 illustrates this process of updating their belief about the session after vocalizing. The marmoset starts with a belief of what call feature elicits the highest response likelihood. After vocalizing and receiving a response (or the lack thereof), this belief can potentially be updated to become narrower, which means more certainty on what is the optimal vocalization to maximize the response likelihood in this session. One way to simulate that process is via using Bayesian update rules, applying the Bayes rule to the belief of the optimal vocalization given a new sampled vocalization and given a variable that represents the presence or absence of a response (see Methods).

Not only different vocalizations have different response likelihoods, but with the Bayesian model we can see that different vocalizations lead to different updates. For example, some vocalizations might not give a lot of new information, and therefore are not very useful for updating the belief. This observation raises the question of whether a marmoset will vocalize in a way that directly maximizes the response likelihood (the peak of the belief), as it is often thought in the theory of vocal accommodation, or one that maximizes the amount of information gained to the belief. Figure 2a and 2b shows these two different policies, which we are called the active sensing policy, and the vocal accommodation policy. The policy is the rule that the marmoset uses to pick a vocalization given the information it already has on their environment. The policy in Figure 2a represents an active sensing mechanism because it is using the agent's energy (their vocalization) in a way that maximizes the amount of new information gained, or minimizes the uncertainty about that session. Notice that by vocalizing at the peak or at the bottom of the distribution, the range of expected response likelihoods are smaller than by vocalizing at the most inclined area of the distribution. This is an intuitive way to understand why these two models would generate different vocalizations, even though it is not true that the active sensing policy will always pick the vocalization in the highest slope of the distribution. A third possibility is that the marmoset already had the information about the session minimized through other mechanisms (i.e. previous information, other senses), and therefore it is not updating the belief regardless of what the new vocalization suggests.
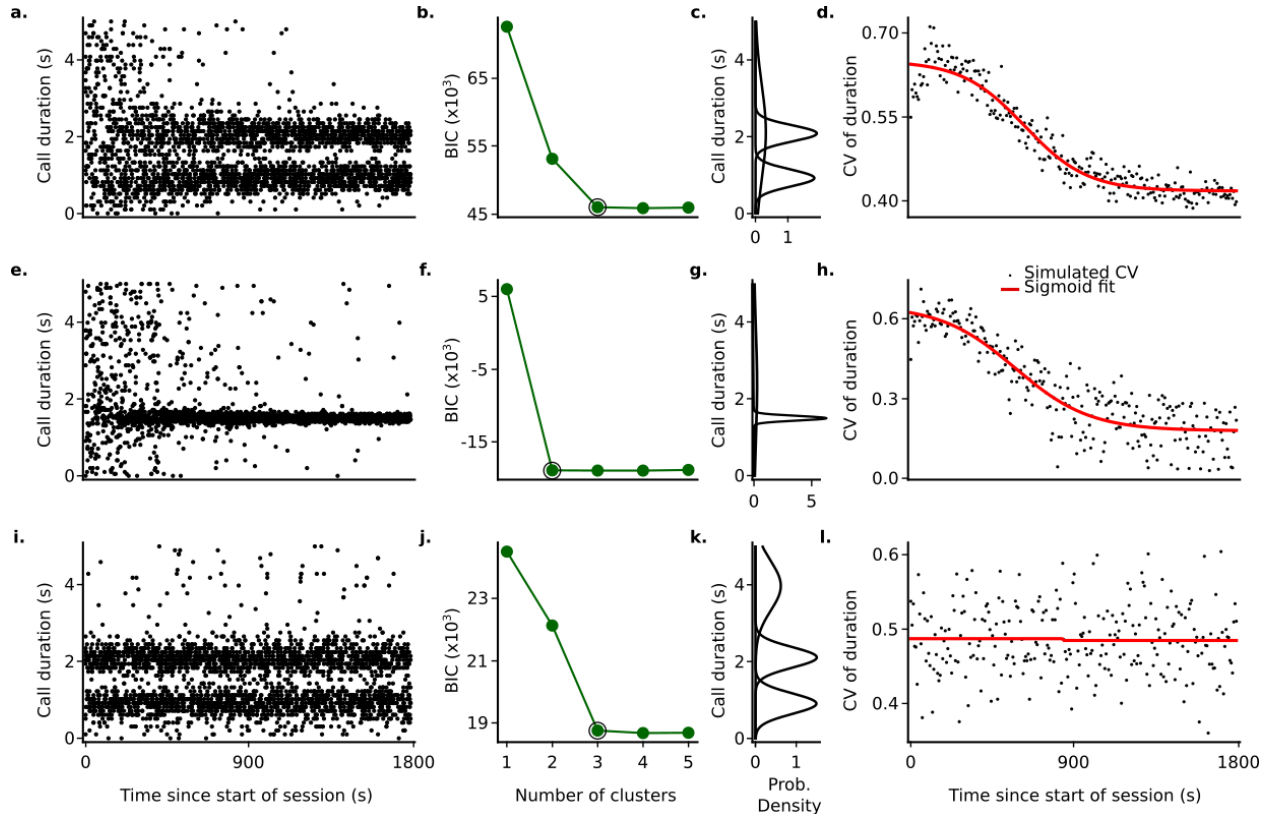
3. The active sensing model predicts a sudden transition and 3 clusters. a. Scatter plot of the simulated call durations in an interaction between two agents using an active sensing model with active sensing policy. **b.** Value of Bayesian information criterion (BIC) versus the number of components when clustering the later half of the active sensing simulation (after 1000 s) with a Gaussian mixture model. The best number of clusters was defined via elbow method and is shown with the dark circumference. **c.** Probability density function of each of the clusters from the active sensing model, using the best number of clusters. **d.** Dynamics of the time-binned coefficient of variation (CV, standard deviation divided by the mean) of the active sensing simulation, calculated via splitting the data into bins in time, and calculating the CV in each bin. Solid Line represents a sigmoid fit of the simulation. **e.** Scatterplot of the simulated call durations in an interaction between two agents using a social sensing model with accommodation policy. **f.** BIC vs number of components when clustering the later half of the accommodation simulation (after 1000 s) with a Gaussian mixture model. **g.** Probability density function of each of the clusters from the accommodation model, using the best number of clusters. **h.** Dynamics of the time-binned CV of the accommodation simulation, along with a sigmoid fit of the simulation. **i.** Scatterplot of the simulated call durations using a no-update policy. **j.** BIC vs number of components when clustering the later half of the no-update simulation (after 1000 s) with a Gaussian mixture model. **k.** Probability density function of each of the clusters from the no-update model, using the best number of clusters. **l.** Dynamics of the time-binned CV of the no-update simulation, along with a sigmoid fit of the simulation.

It is possible to implement these policies with agents that are interacting with another agent that adopts the same policy, i.e., simulate how two marmosets would interact if both were using an active sensing policy for example. In figure 3, each row is the result

of implementing one of the three policies: active sensing, vocal accommodation, and a no-update policy. Figure 3(a-c) shows that the active sensing policy converges from a broad distribution to two main peaks and a third, broader peak, while the vocal accommodation policy (Figure 3(a-g)) converges to mainly one big peak along with a second, broader one. Figure 3i is just repeating the same distribution over time, which was extracted from the active sensing policy since the marmoset has already minimized the uncertainty about the session using other mechanisms, so we already expected Figure 3(j-k) to be similar to Figure 3(b-c). The elbow method of cluster analysis was done in Figures 3(b,f,j) to algorithmically determine the number of clusters. The broad clusters contain the vocalizations outside the main peaks, and are important for the initial wider exploration. Intuitively, the agent tries to explore an area around the vocalization with highest change of response, which allows them to narrow their belief about the optimal vocalization. Figure 3d and 3h shows that both the active sensing and the vocal accommodation policies can exhibit a phase transition in how variable the vocalization is throughout a session, which is not replicated in figure 3l for the no-update policy.

It is possible to see in Figure 4 that the behavior of the marmoset closely matches the active sensing model. In Figure 4(b-c), we show a number of clusters similar to what was observed in Figures 3b and 3j (for the active sensing and no-update policies), while in Figure 4d we observe a phase transition similar to the one observed in Figures 3d and 3h (active sensing and accommodation). This suggests that the active sensing is the best model to concomitantly describe both the dynamics and the clustering, among the 3 models tested. Eliminating the no-update policy means that the marmoset is using their vocalization to update their belief about characteristics of that session, such as the distance of the other marmoset. Eliminating the accommodation policy in favor of the active sensing policy means that this update is not just passive, but something that is being actively pursued by the marmoset.

The first row of Figure 4 has marmosets in a context where they cannot see each other (Figure 1a). If the two marmosets were able to see each other, we expect that most information that was being obtained by the vocalization can now be obtained with their vision (data from Liao et al., 2018). If the vocalizations are not being used with that function, the no-update policy should describe the behavior better. This is what we observe in Figure 4(e-h). Both the clustering (Figure 4(f-g)) and the dynamics (Figure 4h) are better described by the simulations of the no-update policy in Figures 3(i-l).
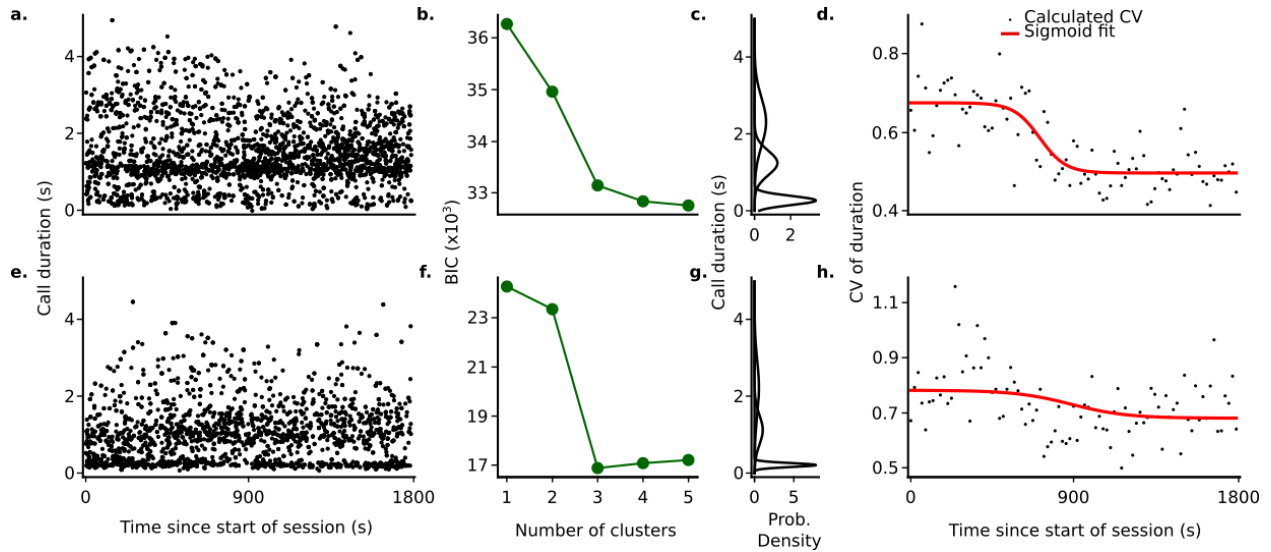
4. Marmosets behavior matches predictions from active sensing model in occluded context. a. Scatter plot of the call duration for each vocalization as it changes through a recording session in an occluded condition. **b.** Value of Bayesian information criterion (BIC) versus the number of components when clustering the later half of the data (after 900 s) with a Gaussian mixture model. The best number of clusters was defined via elbow method and is shown with the dark circumference. **c.** Probability density function of each of the clusters from the data in the occluded condition, using the best number of clusters. **d.** Dynamics of the coefficient of variation (CV, standard deviation divided by the mean) of the data, calculated via splitting the data into bins in time, and taking the CV in each bin. Solid line represents a sigmoid fit of the data. **e.** Scatter plot of the call duration for each vocalization as it changes through a recording session in the visible condition. **f.** Value of BIC versus the number of components when clustering the later half of the data (after 900 s) in visible condition with a Gaussian mixture model. **g.** Probability density function of each of the clusters from the data in the visible condition, using the best number of clusters. **h.** Dynamics of the CV (coefficient of variation), calculated via splitting the data into bins in time, and taking the CV in each bin. Solid line represents a sigmoid fit of the data.

Discussion

Vocal communication is a critical aspect of animal behavior. The manner in which animals, including humans, decide about how to vocalize can provide valuable insights into their cognitive processes and social interactions. In this study, we investigated three potential strategies employed by common marmosets, a species of New World monkey, when choosing how to vocalize. Our results showed that marmosets primarily use active sensing rather than vocal accommodation to guide their vocal dynamics in a visually occluded short-term vocal interaction. This finding represents a departure from previous research on vocal communication in which the vocalizations were important for establishing the most effective vocal exchange in terms of having higher chance of contingent response, such as turn-taking (Takahashi et al., 2016). In the

10

accommodation policy, vocalizations quickly converge to the optimal vocalization with respect to higher chance of response, whereas in the active sensing, vocalizations oscillate around that optimal vocalization with a set of quasi-optimal vocalizations. By adopting an active sensing strategy, the individual is minimizing the uncertainty, since it is choosing the vocalization that minimizes the range of their belief about the environment.

It is important to note that the active sensing policy is not exclusive with vocal accommodation in a broader sense. The correlations between call and response found in Figure 1 show that exhibiting social closeness, as predicted by the social accommodation framework, is still a relevant mechanism in vocal exchange. The differential with active sensing is that it allows for more flexibility in a rapidly changing environment. As it oscillates around the optimal vocalization, when such optimal vocalization changes (e.g. in the presence of another animal, or when the marmoset vocalizing changes their position), the agent is already primed to probe the areas around what it believed to be optimal, instead of insisting on the same vocalization. In other words, it can lead to quicker accommodation, when the environment changes. This explains why active sensing is observed in short term interactions even though accommodation is still observed in the span of days and months.

By improving the speed at which quasi-optimal vocalizations are reached, marmosets are effectively reducing the energy expenditure of interacting with other marmosets, since it quickly avoids vocalizations that are not going to lead to interactions, since each vocalization requires and are constrained by metabolic energy consumption (Zhang & Ghazanfar, 2018). We know that costly-efficient vocalization is also observed in humans, as it is shown by the emergence of Zipf's law (Zipf, 1949 Human behavior and the principle of least effort. (apa.org)). The idea of a cost efficient utterance is also explored when talking about the avoidance of redundancy in human language (Rubio-Fernandez et al., 2021 Least effort and the origins of scaling in human language (pnas.org)). In this context, active sensing can also be a way of avoiding energy waste by efficiently transmitting a signal that will prompt a receiver to provide the speaker with more information.

Other policies for choosing vocalizations still need to be studied. Even though we show that calls can be used that way in an out-of-sight context, that does not inform us about how vocalizations are chosen when marmosets are interacting within their visual field. The no-update policy, as it was implemented here, only informs us that the information from the previous vocalization was not used to update the belief about the environment. We used the information from the active sensing model to simulate that information

being obtained immediately by the visual field. That means that there might be other factors influencing the choice of vocalization that are overlooked by this model, as the model itself does not predict what vocalizations are chosen by relying on the other model.

Methods

Dataset

The dataset reported here was reported previously ([Liao et al., 2018](#)). Recordings were obtained from 6 marmosets splitted in 3 pairs of cagemates, each pair being one female and one male. The marmosets had ad libitum access to water and were fed with commercial marmoset diet, fresh fruits, vegetables, and insects. All experiments were performed with the approval of the Princeton University Institutional Animal Care and Use Committee. We used the recordings from two different conditions: in the occluded condition, in which the pair of marmosets were placed with an opaque curtain between them, and in the visible condition, without the opaque curtain.

Random shuffling and within-session shuffling of the responses

We identified whether a vocalization was a response of a previous call if that supposed response started no more than 12 s after the offset of another call from the other marmoset. For the "original" condition, we merely calculated and bootstrapped the correlation of a specific acoustic feature in a call and its response. We did it for 3 different acoustic features: call type, amplitude, and duration.

We investigated whether the correlation existed only because all vocalizations are similar, instead of being a characteristic of the pair call-response, by shuffling all the responses in that dataset. For each acoustic feature (call type, amplitude, and duration), we had an array for calls, and an array for responses, e.g. call_durations and response_durations, that were aligned so that the i-th element of the response_durations corresponded to the response of the call of the i-th element of call_durations. For each bootstrap in the randomly shuffled condition, we shuffled the elements in the response_durations before calculating the correlation.

Finally, we investigated whether the correlation was present in the session, instead of being a characteristic of the pair call-response, by shuffling responses only emitted by a specific marmoset within each session. In this case, instead of shuffling the whole response_durations array, we'd identify the vocalizations in that array that were emitted in the same session, and then shuffle only the responses within that session. The

shuffle was only complete after each marmoset in all of the sessions had their responses shuffled.

Sensing and observation model

Given that different vocalizations elicit different response likelihoods, the agent (a marmoset) wants to know what is the optimal call duration θ to produce to get the highest response likelihood V. The estimated θ will be tracked via the belief $p(\theta)$ of what is the optimal vocalization.

Let's define a function $\Upsilon(\theta, x)$ that, given a sampled call duration x and the optimal call duration θ, returns the average response likelihood to that vocalization, so $V = \Upsilon(\theta, x) + \epsilon$, where ϵ is an observation noise. That function can be modeled as the following Gaussian:

$$\Upsilon(\theta, x) = \exp\left[-\left(\frac{x-\theta}{2\sigma_1}\right)^2\right]$$

so that the minimum response likelihood is 0 when x is very different than θ, and the response likelihood is 1 otherwise. The parameter σ_1 in the function is how broad that function is, still peaked at 1, which means that a big σ_1 makes it easy for x to be associated with a high response likelihood.

Bayesian update of the belief

We followed the same modeling strategy from [(Chen et al., 2020)](#), using a Bayesian filter to update the belief. However, in their case, after taking an action, the agent receives a continuous number as a measurement from their action. In our case, after taking an action (vocalizing), the agent receives a discrete measurement - whether there is a response or not - and use that measurement to estimate the continuous value underlying that response (the response likelihood V).

Let's say that $r \in \{0, 1\}$ is the variable that represents the response, $r = 1$ if there is a response. To updated the belief given the response, we can use the following Bayesian filter instead:

$$p(\theta \mid r, x) = \frac{p(r \mid \theta, x)\ p(\theta)}{p(r \mid x)} = \eta\ p(r \mid \theta)\ p(\theta)$$

Notice however that $p(r \mid \theta, x)$, i.e., the chance of receiving a response given θ and x, is directly related to the observational model $\Upsilon(\theta, x)$, so with η as a normalizing factor, we have

$$p(\theta \mid r, x) = \eta \ \Upsilon(\theta, x) \ p(\theta)$$

Active sensing and the Expected Information Density (EID)
The policy that leads the agent to actively pick the vocalization that is most informative (active sensing), i.e., that leads to the biggest narrowing of the belief, is taken from the expected information density as described in [(Chen et al., 2020)](). Each belief has a Shannon entropy associated, which is calculated via $S[p(\theta)] = -\sum_{\theta} p(\theta) \log p(\theta)$.
The lower the entropy is, the more certainty we have in the belief. Therefore, the best vocalization to sample, is the one that lowers the most the belief after updating, i.e., the one with the lowest $\Delta S = S[p(\theta)] - S[p(\theta|V, x)]$.

Given a prior distribution, the model will simulate what would happen when sampling each value in the whole vocalization space (it will go loop through all values of θ), and for each value of x, it will be estimate what the updated belief would be for for each value of V (another loop through V).

For each V, we calculate the entropy reduction ΔS for each potential vocalization x, which gives a $\Delta S(V, x)$, in function of x. Then, we calculate the measurement probability:
$$p(V|x) = \int_{\theta} p(\theta)p(V \mid \theta, x)d\theta$$
, which we use as the weight of $\Delta S(V, x)$. Finally, we integrate over all possible values of V, so that:
$$\mathrm{EID}(x) = \mathrm{E}\left[\Delta S\right] = \int_{V} p(V \mid x) \ \Delta S(V, x)dV$$

Decision making for no-update, sensing, and active sensing policies
In all of the policies, the call duration chosen for the next vocalization is sampled from a probability distribution. For the no-update policy, the probability distribution is constantly the distribution of all calls durations from the real data, normalized to a probability distribution. For the sensing policy, even though the vocalizations are used for sensing and updating the agent's expectations (the belief), the call duration is sampled from the belief itself, so that the action is taken to maximize the response likelihood in the vocal exchange instead of maximizing their sensing potential. For the active sensing policy, the call duration is chosen from the expected information density (EID), normalized to a probability distribution. That way, the vocalization is chosen so that it maximizes the entropy reduction of the belief, i.e., the sensing potential.

References (not formatted)

1. Zweifel & Hartmann, 2020 Defining "active sensing" through an analysis of sensing energetics: homeoactive and alloactive sensing - PubMed (nih.gov)
2. Jones et al., 2021 Communication with self, friends and foes in active-sensing animals - PubMed (nih.gov)
3. Luo et al., 2021 Constraints on the deformation of the vibrissa within the follicle | PLOS Computational Biology
4. Bajcsy, 1988 Active perception - Proceedings of the IEEE (berkeley.edu)
5. Gibson, 1962, Gibson (1962) Observations on active touch (zhdk.ch)
6. Beleyur, 2019 Modeling active sensing reveals echo detection even in large groups of bats | PNAS
7. Capranica, 1965 https://psycnet.apa.org/record/1965-14349-000
8. Penna, 1997 https://doi.org/10.2307/1565388)
9. MacIver et al, 2010 Energy-Information Trade-Offs between Movement and Sensing | PLOS Computational Biology
10. Chen et al., 2020 Tuning movement for sensing in an uncertain world | eLife (elifesciences.org)
11. Varella et al., 2022
12. Bezerra & Souto, 2008 Structure and Usage of the Vocal Repertoire of Callithrix jacchus | SpringerLink
13. Ruch et al., 2017 The function and mechanism of vocal accommodation in humans and other primates (wiley.com)
14. Takahashi et al., 2016
15. Zhang & Ghazanfar, 2018
16. Zipf, 1949 Human behavior and the principle of least effort. (apa.org)
17. Rubio-Fernandez et al., 2021 Least effort and the origins of scaling in human language (pnas.org)
18. Liao et al., 2018

Brainstorm:

One of the current trends in engineering and computational social sciences is social sensing, an umbrella term being used to describe techniques and devices used to collect human data as they interact socially (e.g. in social media) (Liu et al., 2015 https://doi.org/10.1080/00045608.2015.1018773 ; Wang et al., 2019 https://doi.org/10.1109/MC.2018.2890173). Yet, this technique predates the use of any modern machines. Humans and non-human animals alike are constantly using their sensory organs to gather information about conspecifics, and make inferences on how to behave based on that information. For example, dogs can gather information about another dog's diet and immune system via sniffing (Nara et al., 2011 10.4081/ejh.2011.e29), and marmosets can learn about another individual's gender, health, and arousal (among other things) via listening to their vocalization. Not only they acquire information, but they accumulate that information and act accordingly (accumulation-of-evidence in decision making, e.g. Pinto et al., 2018 https://doi.org/10.3389/fnbeh.2018.00036), and sometimes act in a way that seeks to gather more information. For example, in the weakly electric fish, this process is mediated by a hippocampus-like region to aid them in navigating a new environment (Fotowat et al., 2019 https://doi.org/10.7554/eLife.44119).

Another well described example of active sensing for navigation is the electric fishes use of electrolocation (Hofmann et al., 2017 Sensory Flow as a Basis for a Novel Distance Cue in Freely Behaving Electric Fish (jneurosci.org)). Electric fishes generate electric discharges via their electrocytes, which provides them with information about nearby objects or other animals, helping with navigation and with foraging.

First should be on Social Sensing, broadly speaking. Humans (linguistic laws?). Songbirds. Cite neuro stuff on social navigation (Witten Lab for example, or Cory Miller hippocampus). No one has a handle on the behavior.

Second: models of sensing; elaborate on MacIver's literature

Third is about vocalizations as a standard means of social sensing, particularly out-of-sight:

Fourth: introduce the marmosets for vocal interaction and then model that you are going to use.

DISCUSSION NOTES:
Summary of Results
Interpretation of results
Related to other findings: Optimization of energy/Linguistic laws
Reinterpretation of other marmoset publications. E.g.: marmosets accommodate on a larger temporal scale - why use active sensing? Flexibility and speed - active sensing as the short term implementation of an accommodation.

*environmental accommodation, social accommodation: accommodate to a partner in a short period, signal social closeness The function and mechanism of vocal accommodation in humans and other primates - Ruch - 2018 - Biological Reviews - Wiley Online Library