Marmosets Communication - Unfolding Identity

Guy Paul Hadad, Tomer Fried

October 1st, 2023

Abstract

The ability to replicate and modify sounds by non-human primates was considered almost nonexistent. Recent studies challenge the consensus of vocal learning abilities among non-human primates. To test these new revelations, with regards to the use of identity in communication, we delved into the communication of Common Marmosets. We took calls of individual marmoset monkeys which were recorded between pairs of individuals and believed to contain factors related to identity. We wanted to check the existence of calls which are directed to specific individuals, like the way humans call themselves by names. This hypothesis is split into: Global directed calls – a common call, used by multiple individuals and directed to a specific one. Local directed calls - a call used by a single individual and directed to a specific monkey. Using tools and methods from supervised learning, we discovered significant evidence which implies the existence of local calls, while for global calls we could not find significant evidence.

Introduction

Background

Vocal communication is a common social behavior among many species, in which acoustic signals are transmitted from sender to receiver to convey information such as identity, individual fitness, or the presence of danger. Across diverse fields, a set of shared research questions seek to uncover the complexity and depth of vocal communication of various species. Among non-human primates, there exist many areas of uncertainty regarding vocal communication: How are signals produced and perceived? What information is carried within signals? And specifically, do these signals carry within them information related to the identity of the caller (sender) and callee (receiver)?

<u>Vocal Learning and its relation to identity signals</u>

In the past few years more research has come to suggest that monkeys - a sub-category of primates - have higher capabilities than previously thought with regards to the ability to listen, and then reproduce vocalizations they have heard in a precise manner or with chosen modifications. This ability is known as Vocal Learning. In our research we seek to find out whether Marmoset monkeys can use Vocal Learning in a nuanced way to convey aspects of identity in their vocalizations. And more specifically - can a single individual marmoset (caller) produce a call to a different marmoset (callee) in which the call has aspects of identity referring to the callee? And also, can multiple callers do the same to one specific callee - showing that they have one single and agreed upon way to address each individual in the group?

<u>Connection To Neuroscience</u>

The significance of answering these questions and understanding the characteristics of vocal learning among monkeys can provide insights about the evolution of human language, as the ability to perceive sounds from other monkeys and perform modifications to them is considered a cornerstone for language development. In addition, studying vocal learning among monkeys can help us understand the neurological and cognitive mechanisms behind this ability, and as a result shed light on the same mechanism among humans, who as we know, only undergo such experiments under extreme constraints.

Related Work

Our starting point is based on “Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires” (Sainburg et al., 2020). This paper presents a set of computational methods for projecting animal vocalizations into low dimensional latent representational spaces that are directly learned from the spectrograms of vocal

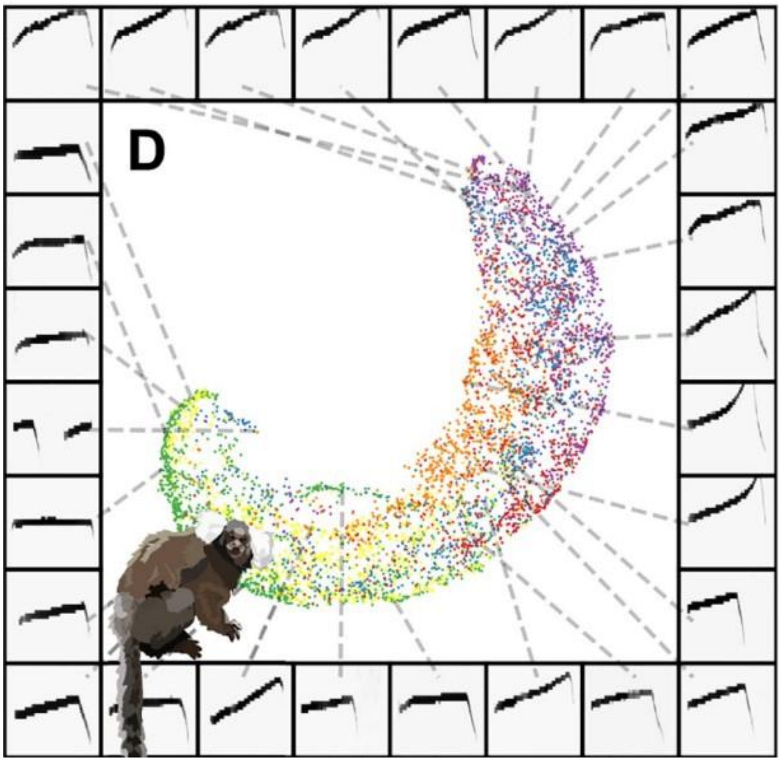

Figure 1: “Calls embedded into a 2D UMAP space, where each point in the scatterplot represents a single element. Scatterplots are colored by individual identity. The borders around each plot are example spectrograms pointing toward different regions of the scatterplot.” (Sainburg et al., 2020)

signals. One of the experiments done in this article was focused on testing the ability to identify a caller by its Phee call. A UMAP latent space representation of spectrograms of different callers’ Phee calls has shown a spectrum which is roughly clustered into individual callers. These results serve as a basis to our question of callee identity within Phee calls.

Data

To answer this question, we have received 45,000 Marmoset vocalizations from 8 individuals, which participate in different social groups. These vocalizations are known as 'Phee Calls', and are perceived in scientific community to contain aspects related to the identity of the caller and callee. To record such calls, Prof' David Omer's laboratory has conducted an experiment, in which pairs of individual monkeys were put together in a room, 2-3 meters away from each other with a curtain between them. The Phee calls were extracted from audio recordings of these sessions and processed into spectrograms - visual representations of the frequencies and amplitudes of a sound signal as it varies over time, and each individual cell represents the strength of a signal at a specific time. The spectrograms were created using Short Time Fourier Transform, under the representation of Mel-frequency cepstrum – which allows to represent the sound's spectrum in a more compact and perceptually meaningful way for the human ear, by focusing on the most important frequencies and how they change over time. The spectrograms are paired with labels that signify the identity of the caller (the monkey that produced the vocalization), the callee (the monkey who was called) and the date of the experiment in which these vocalizations took place.

Results

Prediction of Callees – Full Data

At first, we wanted to see whether by using all the callers, we could achieve a statistically significant prediction of all the callees. To do this, we ran a Random Forest model, which we fit to a train set of the data and predict on a test set. We thought that this could give us a general idea and starting point regarding whether there is similarity between calls from different monkeys to specific individuals.

We analyze the results of the predictions by looking at a confusion matrix, which averages 100 confusion matrices of different predictions of randomly created train-test sets of data, whether by train-test split of the data, or by In-Bag train set and Out-Of-Bag test set (OOB). The y-axis represents the true callee and the x-axis represents the predicted callee of the calls.

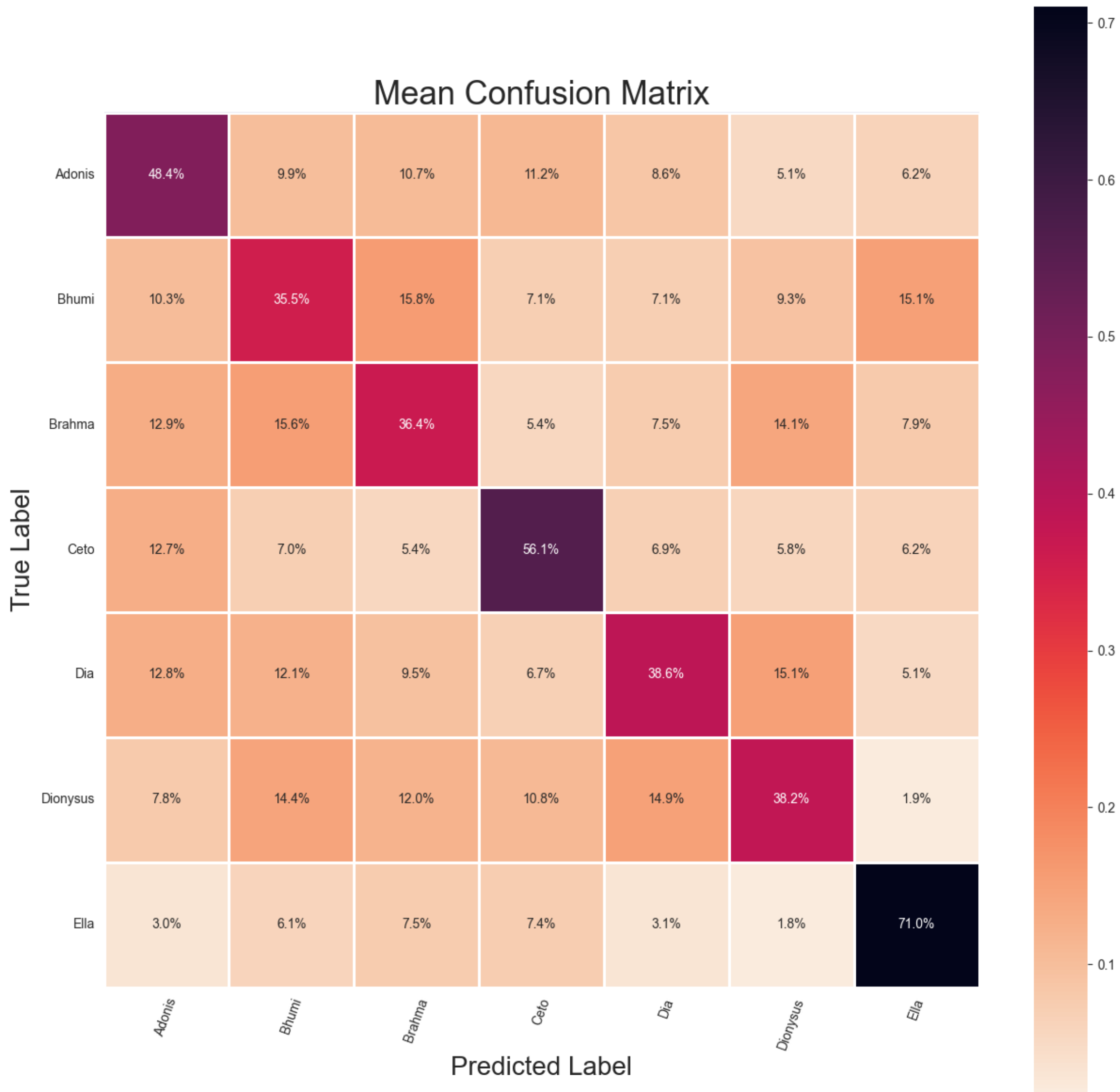

Figure 2: Mean confusion matrix, OOB test set

The mean confusion matrix shows for every individual a significant number of valid predictions of a specific callee: The diagonal of the confusion matrix represents samples in the test set which their predicted callee matches the true callee. The percentage of successful predictions, out of all predictions per callee, is above chance level (14.28%) for every individual, with a range of 37.5% - 71% success rate.

Prediction of Callees – Per Caller

Since the previous prediction showed some interesting results, we wanted to check if a similar pattern of prediction exists when the train and the test data belong to a single caller only

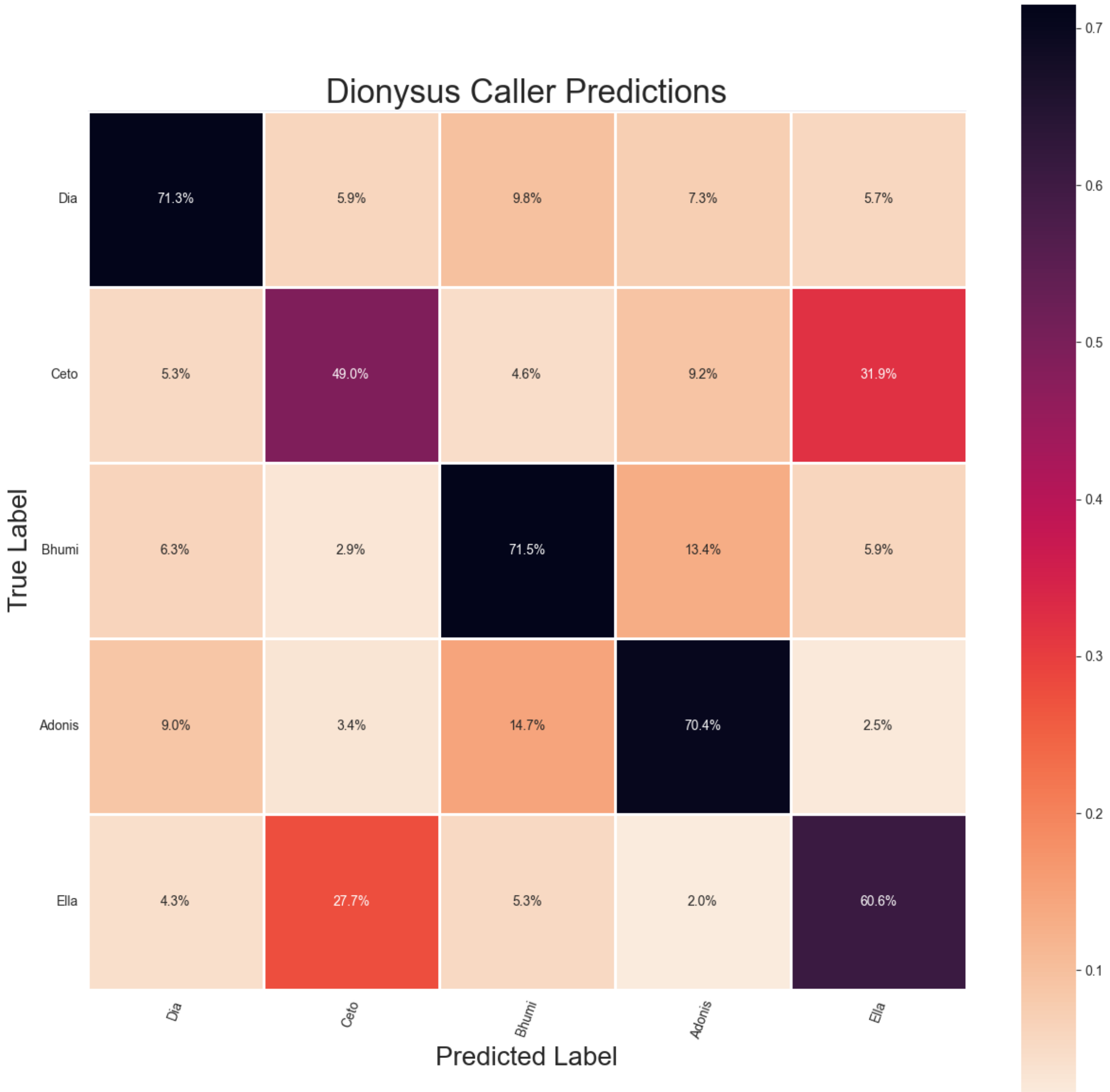

Figure 3: Mean confusion, Dionysus is caller in the data, OOB test set

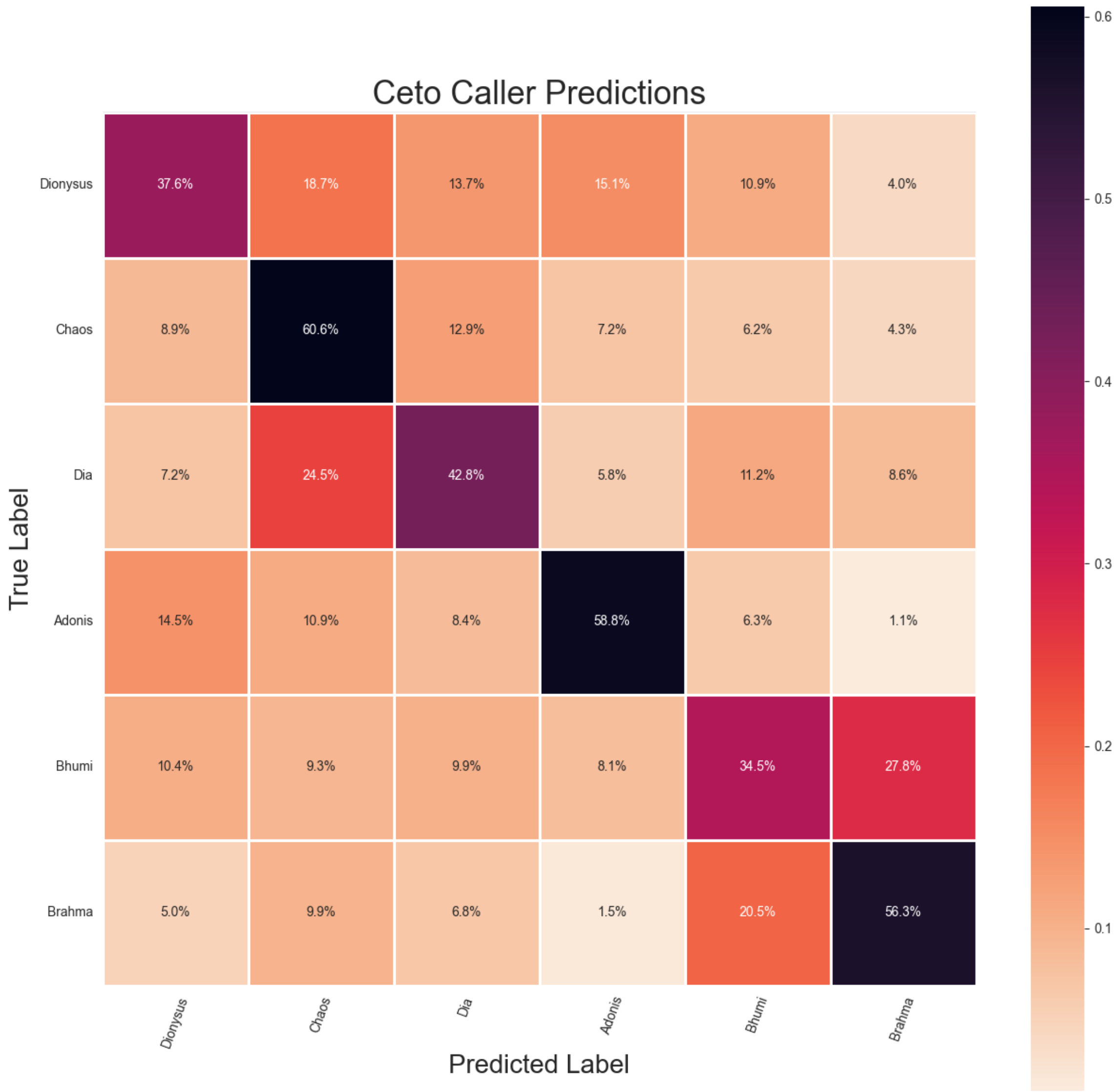

Figure *4*: Mean confusion, Ceto is caller in the data, OOB test set

From the results we can observe that the same pattern of prediction exists when using the per-caller train data as well. Also, some confusion matrixes have an even higher percentage in diagonal elements of the confusion matrix than seen before.

Prediction of Callees – Train on a Caller Test on The Rest

We wanted to see if there's a significant prediction when the model is trained on a single caller's data, and then predicts on the rest of the data, to see if there's evidence for a global identity signal for individuals, which is identical for multiple callers, and is related to a callee.

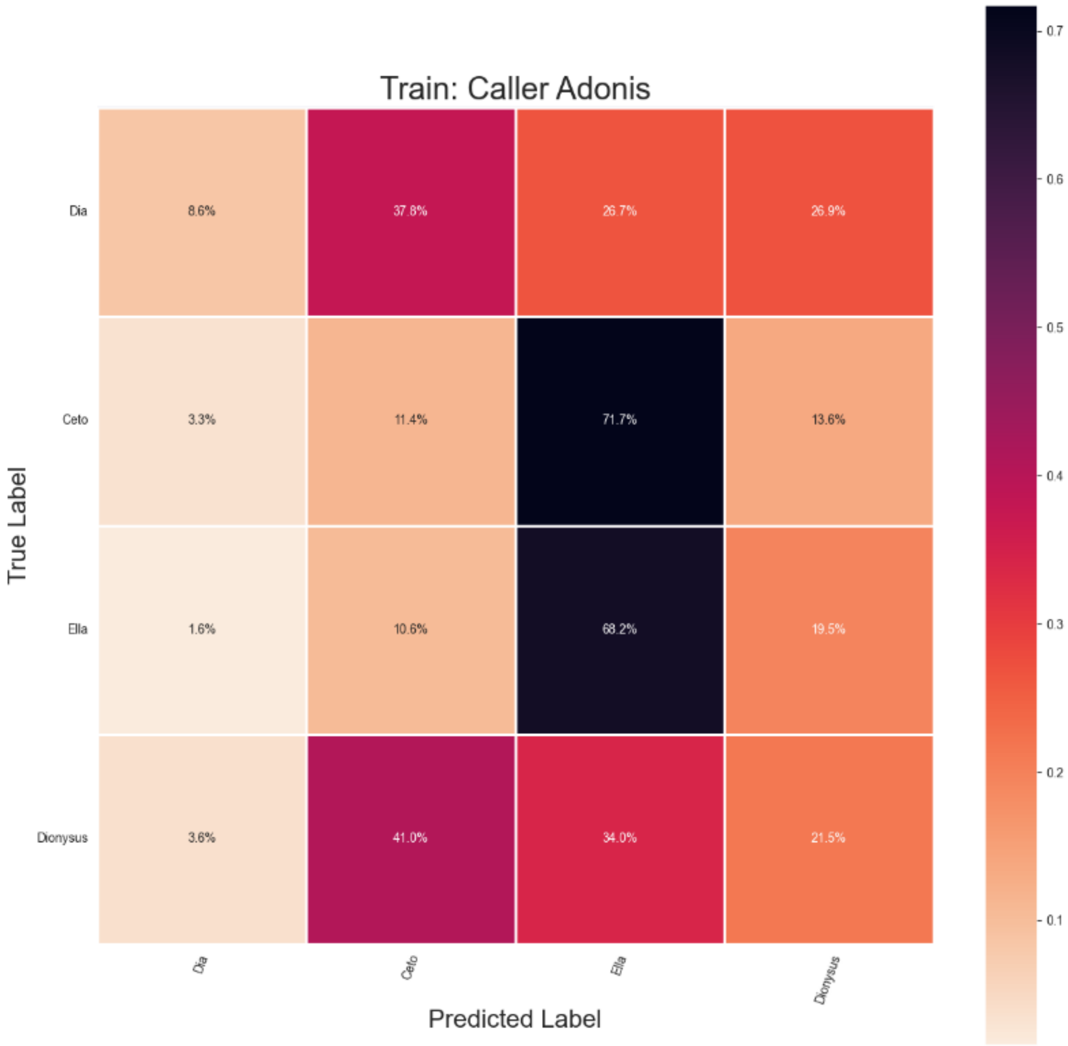

Figure 5: Mean confusion matrix, caller is Adonis in the train set

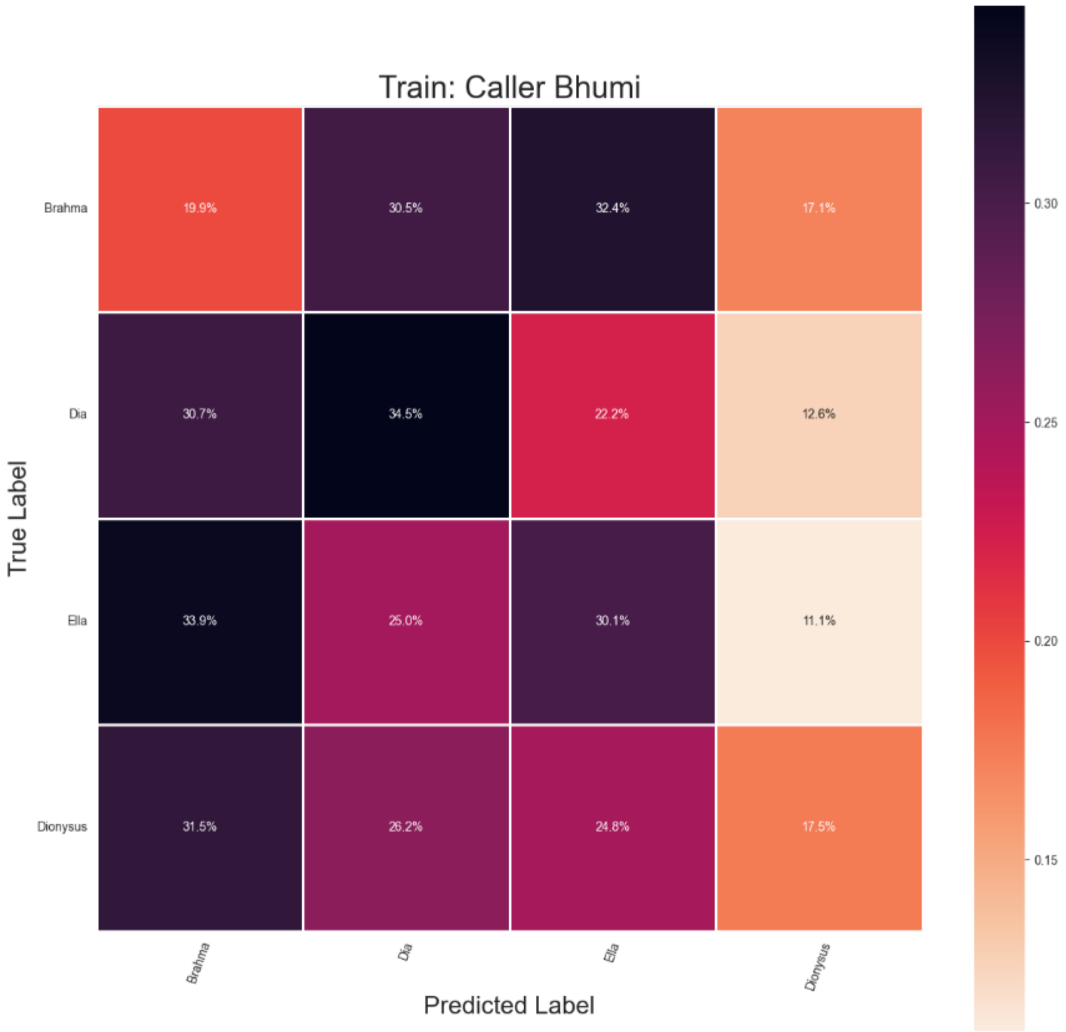

Figure 6: Mean confusion matrix, caller is Bhumi in the train set

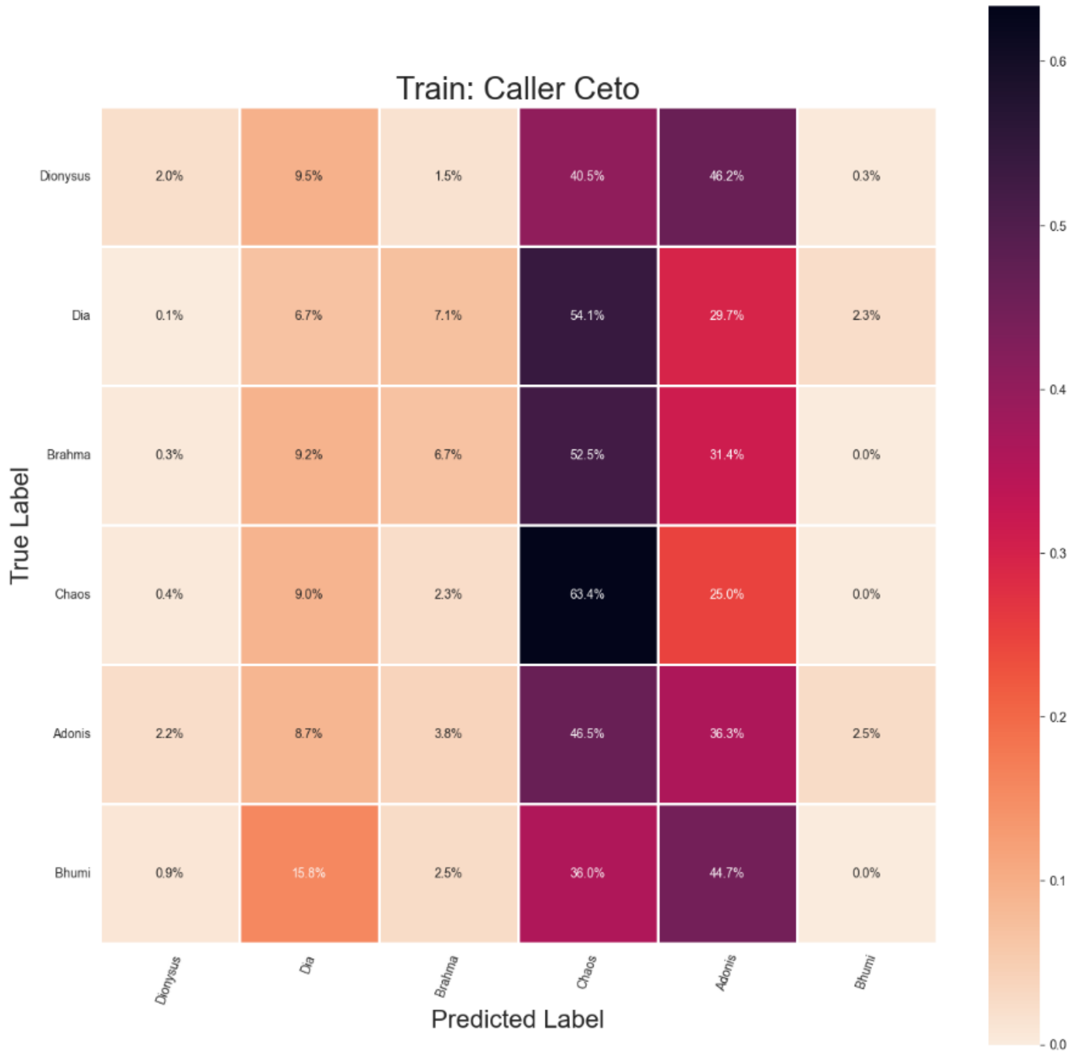

Figure 7: Mean confusion matrix, caller is Ceto in the train set

We did not receive significant predictions. The confusion matrices do not produce a diagonal with an above chance level percentage of success in most diagonal cells of the confusion matrices. In addition, we can see that some of the confusion matrices columns have higher percentages than others.

Predictions By Selected Features

Looking at the spectrograms representing the calls, we wanted to try and reduce noise that might have come from different experiments. To do this, we thought of representing each call by the following:

- Maximal frequency level per time unit
- Value of maximal frequency level per time unit (amplitude)

Experimenting with those features as our data, we found out that we receive similar results to previous experiments with the spectrograms as our data.

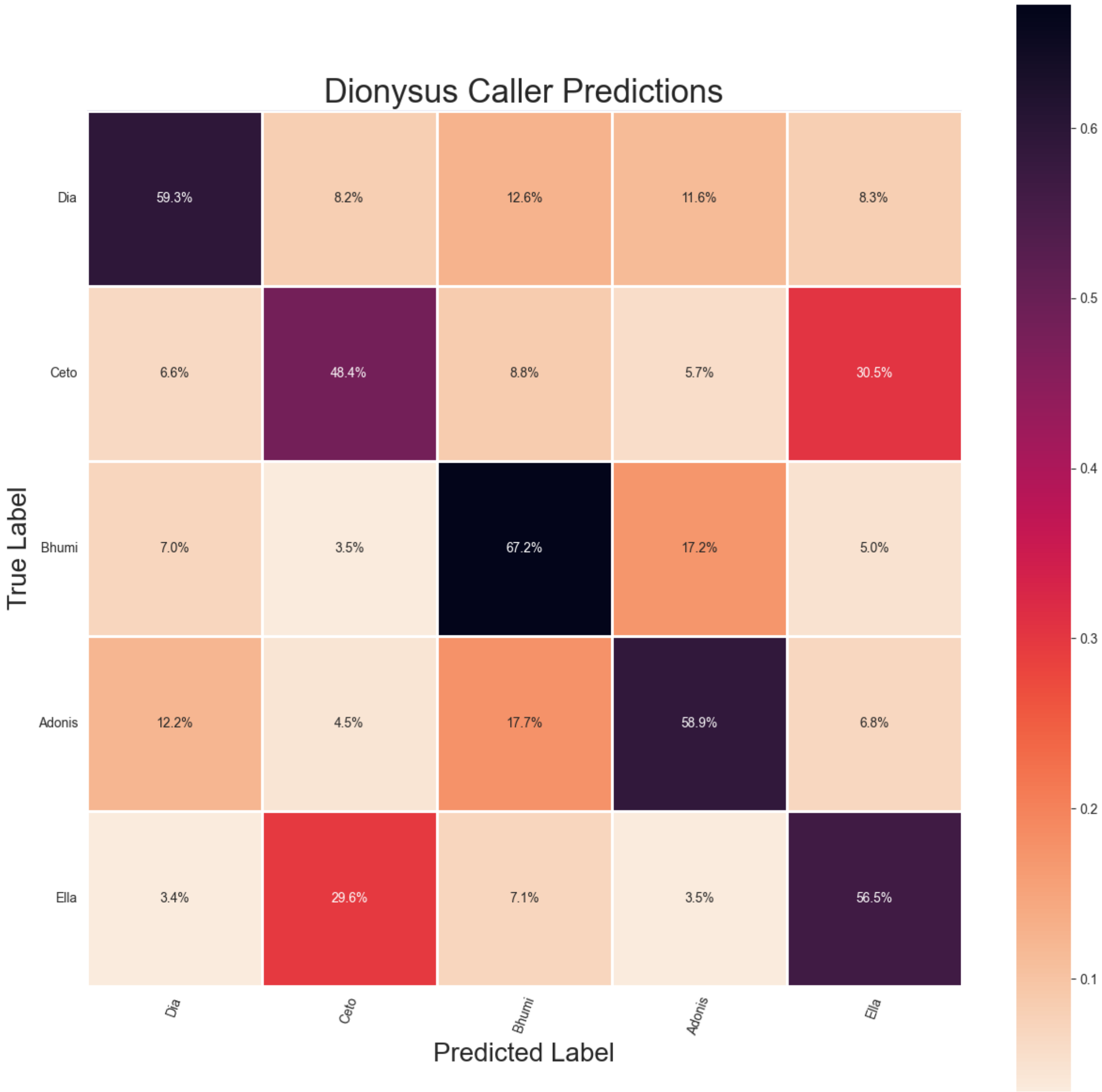

Figure 8: Mean confusion, Dionysus is caller in the data, OOB test set

Link to previous prediction with spectrograms

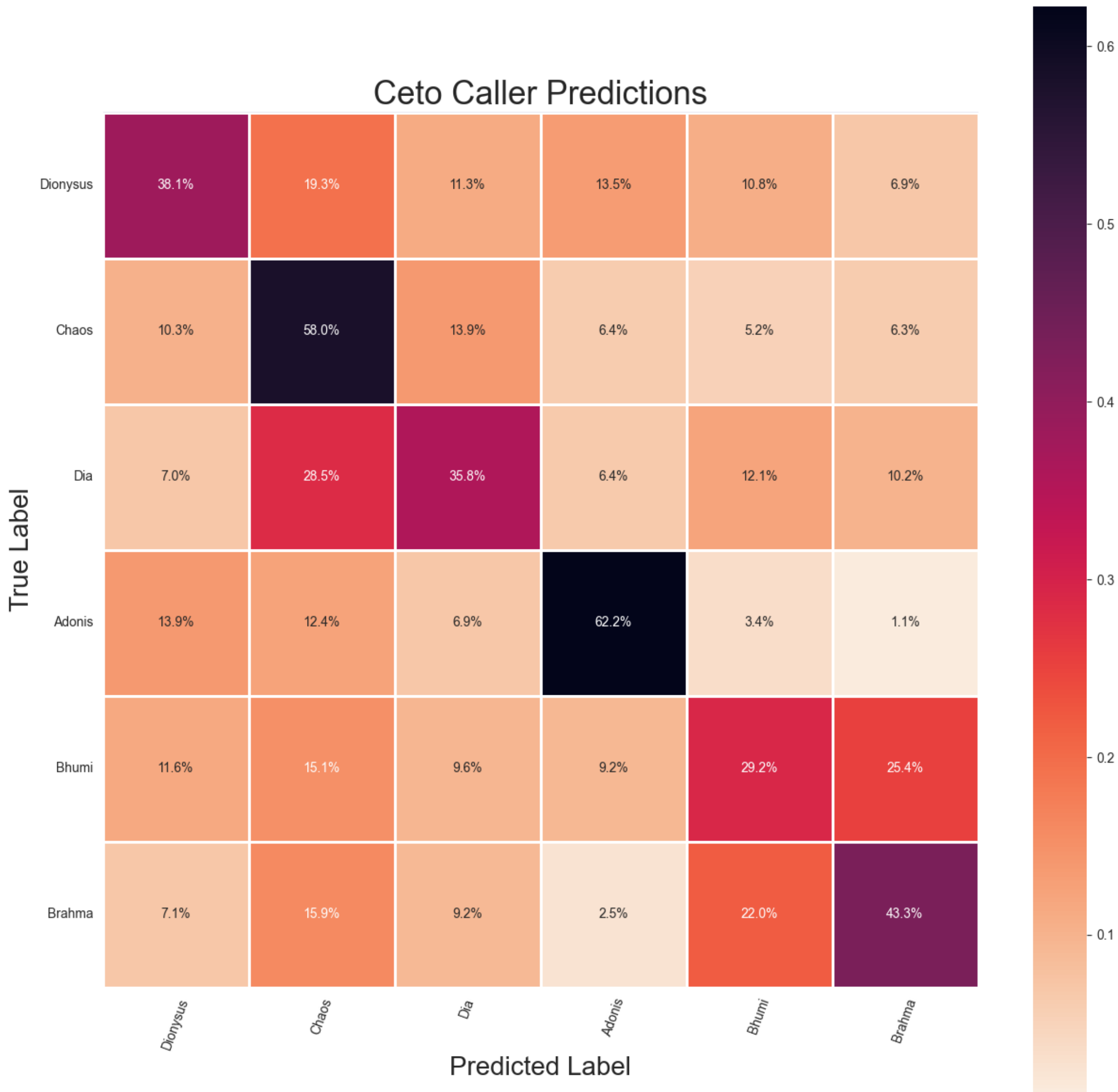

Figure 9: Mean confusion matrix according to the, Ceto is caller in the data, OOB test set

[Link to previous prediction with spectrograms](#)

Discussion

Results Summary

From the experiments we found the following:

- We were able to predict callee identity based on multiple caller dataset with above chance level of certainty.
- We were able to predict callee identity based on a single caller dataset with above chance level of certainty.
- We were not able to predict callee identity when trained data is of a specific caller and tested data is of other callers.
- The above-mentioned results were achieved either on a dataset of spectrograms of calls, or a dataset of maximal frequency per time unit and their respective amplitudes.

Conclusions

The above results indicate:

1. A single individual marmoset (caller) can produce a call to a different marmoset (callee) in which the call has aspects of identity referring to the callee.
2. Features of maximal frequency per time unit and their respective amplitudes which are extracted from the spectrograms produce similar results to using full spectrograms. These results indicate that these features may be a good substitute for the full spectrograms, because they might reduce noise and are of lower dimension.
3. We did not find any significant results which prove an existence of global identity calls – meaning that there was a lack of evidence for multiple callers which refer to one specific callee with a single and agreed upon way.

Limitations

Although the experiments and research were conducted thoroughly and met the required standards, they did have a few limitations that are worth considering:

- The experiments were performed using pairs of individuals, separated from the others. It was assumed that interactions of isolated pairs are reflective of normal interaction. but in actuality, it might not represent the natural interactions between the monkeys, thus implying that we need to treat the results with less confidence, if true. The experiment could have been performed better if the calls were collected in a more natural and organic setting, such as their regular positioning in the laboratory where they are all together. An even better solution would be to record them in nature. Even though these suggestions are better, they are much more difficult to implement in reality.
- Our predictions were all produced by Random Forest ensemble machine learning model. Even though we have received significant results using this model, it might have been better to use a model that can identify more complex structures and connections in the data, and thereby might affect the percentage of right predictions.
- In addition, due to a possible drift along time, it might be necessary to record all the calls in a smaller timeframe – At most a few months. Some of the calls were recorded during a time frame of 1-2 years, in which we have spotted differences in the calls over time, which potentially might have harmed the percentage of the right predictions. This change was visible using [Autocorrelation].

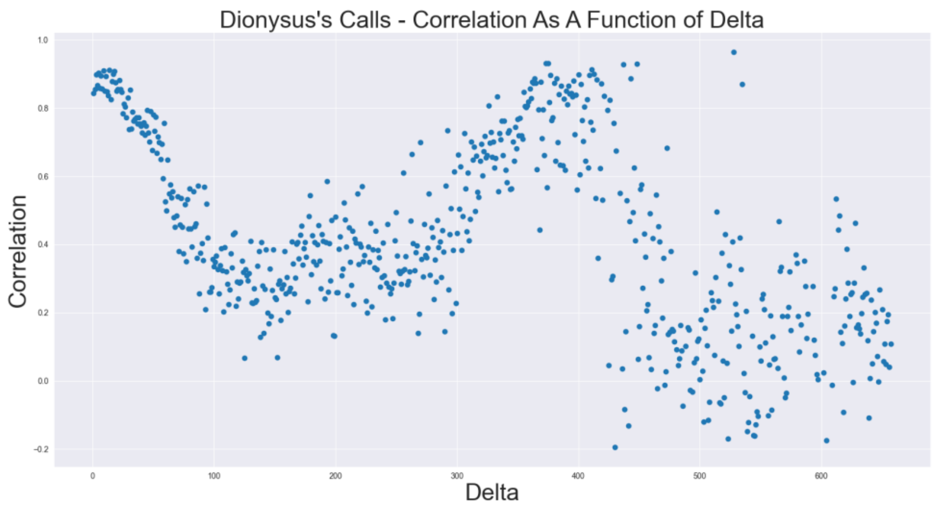

Figure 10: X-axis is the difference between session dates in days, Y axis is Pearson correlation.

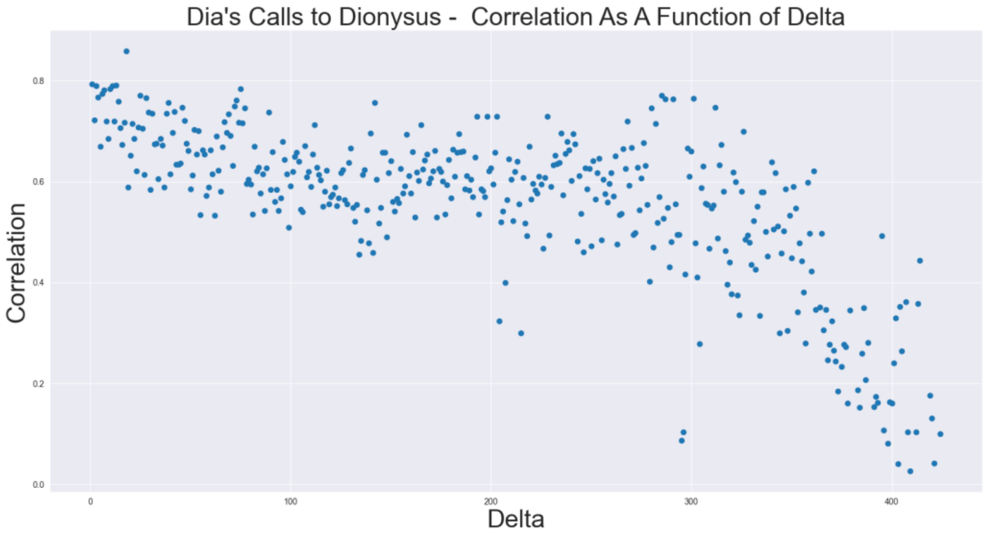

Figure *11*: *X-axis is difference between session dates in days, Y axis is Pearson correlation.*

Future Work

Our work mostly shows that it is possible to predict callees with above chance level certainty, based on a specific caller. This means that the identity signal seems to rely uniquely within pairs of individuals. On the other hand, an identity signal which is used by multiple callers and is directed at a single callee is not apparent in our results. Based on these findings, we believe that future work should focus on:

- Since marmosets form different social groups and families, it should be further investigated whether specific groups that we had not tested for, practice global-identity calls. This direction of research was raised when we conducted the "Train on a Caller Test on The Rest" prediction, where we could spot a trend where the model seems to classify in the test set the partner of the caller in the train set, even though the amount of calls to unique callees in both train and test set were equally balanced. This may indicate that groups of unrelated monkeys learn how to call each other by hearing partners calling each other.
- Due to the model having a limited expression ability, it would be advised to use more expressive models, which could help us find complex relations in the Phee calls.

- It might be helpful to refer to different research papers, especially of closely related species, which show similar results to what we have found. Such articles might be helpful to understand the role of identity in marmosets' communication.

Resources and Methods

Preprocessing

Along with the project we used multiple processing methods on our data:

Feature Extraction – Initially, we used the spectrograms flattened into vectors as our features. During our research, since noise in the spectrograms interfered with our predictions, we decided to remove it by extracting the following features:

- Frequency of Maximal signal Per time bin
- Amplitude - Value of Frequency of Maximal signal Per time bin

Drop entries which their target label has low occurrence in the data – In cases of train sets with extremely low entries per target label compared to the others, we chose to drop the target label of the minority group, to avoid unreliable results.

Under-Sampling – Under-Sampling a dataset such that the number of entries per unique target label is balanced according to the minority label count of entries. This method helped us to reduce bias in the Random Forest model, which is affected by the distribution of different target labels in the dataset. We did this since this distribution was formed due to the way calls were recorded, and not due to a natural phenomenon.

Statistical Methods and Considerations

Cross Validation – To reduce variance in our confusion matrices, we chose to run 100 iterations of the Random Forest model, each consists of randomly created train and test sets. Each model was then fit on a train set and predicted on a test set. Finally, all 100 confusion matrices were averaged into a single final one, which is row normalized.

Out Of Bag Score – During our research we had to deal with small datasets. To avoid dropping critical entries by the Train Test Split method, we chose the test set to be the OOB samples.

Autocorrelation – Since our session's predictions have indicated a possibility of significant change of an individual's Phee calls through time, we wanted to check this hypothesis. The process was inducted as such:

- Every spectrogram is flattened into a vector.
- For every unique date of sessions, take the flattened spectrograms, and calculate an element-wise mean between them.
- The spectrogram which has the smallest distance to the mean spectrogram is considered as a representative of this date of sessions. The distance is calculated between the mean spectrogram and all relevant spectrograms, based on the Euclidean distance between the two.
- Between every two representing session spectrograms, we first calculate the Pearson correlation coefficient, and then pair it with the time difference (in days) between both sessions which we call 'delta'.
- An average is calculated between correlation coefficients of sessions with the same delta, leaving each delta with a single averaged correlation coefficient scalar.
- Finally, the correlation coefficients are shown on a scatter plot, when the x-axis is the delta, and the y-axis is the correlation coefficient.

Medoids Comparison – Since our predictions of the Random Forest were not interpretable due to the nature of our features, we wanted to eyeball the spectrograms, in hope that we might intuitively be able to pick up on factors that are involved in the prediction / indicate which monkey is called. Since there are tens of thousands of spectrograms, we wanted to look at a Medoid spectrogram, which represents a subset of data:

- We took a subset of the data and calculated the element-wise average spectrogram.
- We then calculated the closest spectrogram to the average one, based on the Frobenius norm of their difference, to finally receive the medoid.

Feature Importance Matrix – To understand which features contributed the most for the prediction, we utilize the feature importance tool that "Sci-kit Learn" Random Forest has, which indicates each feature's importance by a value between 0-1 (0-> unimportant, 1->important), where each cell of the spectrogram is a feature.

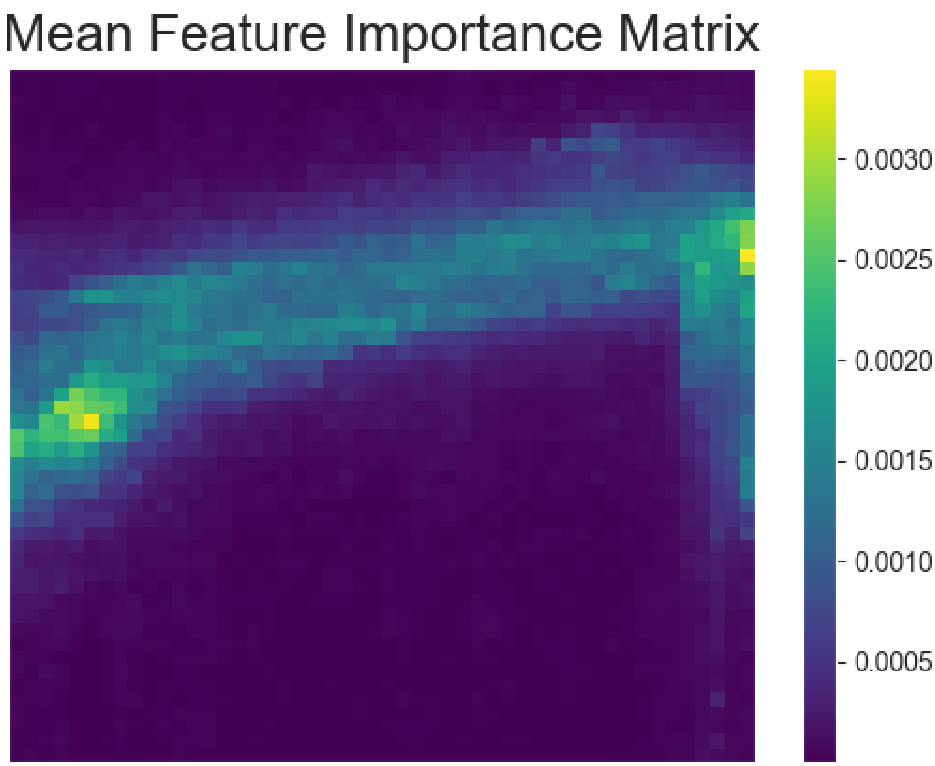

Figure *12*: Visualizes the importance of every element in the spectrogram to the prediction

Models and Machine Learning Methods Used

Initially, we chose to use Random Forest as our baseline model. During our research, we found that it produces good results. Since our main focus in the project was to achieve above chance level prediction, which was accomplished already, David and us decided to continue our work with it.

Work Pipeline

Along with the research process, our work pipeline was as follows:

- <u>Discussion</u> - Meeting with Prof. David Omer: Looking at the previous two weeks' results, discussing and making conclusions based on them, and accordingly setting tasks and implementations to pursue.
- <u>Design</u> - collaborate to develop a detailed plan for how to carry out the tasks and objectives that were set. This involves identifying the resources and tools

needed, defining the roles and responsibilities of each partner, and establishing a timeline for completion.

- Execution - actively carrying out the tasks and objectives that were set based on the conclusions drawn from the previous two weeks' results.
- Review and Preparation - During this phase, we evaluated the results of the execution phase and discussed their findings. We analyzed the data and outcomes to determine whether the objectives were met and to identify any areas for improvement. We then prepared a summary of the results and conclusions to present to Professor David Omer. This involves organizing the information in a clear and concise manner and may include creating visual aids such as charts or graphs to help illustrate the findings.

References

Tim Sainburg, M. T (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. page 10.

OOB - Gareth James, D. W. (2023). An Introduction to Statistical Learning. page 345.

UMAP - Leland McInnes, J. H. (n.d.). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.

Random Forest - Gareth James, D. W. (2023). An Introduction to Statistical Learning. page 346.

Feature Imp*ortanc*e - Sci-kit Learn - Feature importances with a forest of trees