DECISION OF LEARNING STATUS BASED ON MODELING OF THE INFORMATION MEASUREMENT OF SOCIAL BEHAVIORAL TASKS IN RHESUS MONKEYS

Seunghyun Lee

Department of Electrical and
Computer Engineering
The University of Arizona
1230 E. Speedway Blvd.
Tucson, AZ, USA

eacham@email.arizona.edu

Katalin M. Gothard

Department of Physiology
College of Medicine
The University of Arizona
1501 N. Campbell Ave.
Tucson, AZ, USA

kgothard@email.arizona.edu

Jerzy W. Rozenblit

Department of Electrical and
Computer Engineering
The University of Arizona
1230 E. Speedway Blvd.
Tucson, AZ, USA

jerzyr@email.arizona.edu

ABSTRACT

We are interested in identifying the learning status of the social behavioral tasks in the rhesus monkey. In addition, we define the characteristic of stimulus with a numerical quantification. We allow monkeys to interact with individuals of different social status, while we monitor the viewer monkey's behavior by tracking its scan paths. With these observations, we can understand the learning status of this animal via looking behavior analysis on the stimulus. First, the viewer monkey shows different looking patterns among six different classes. Therefore, we can generate different data descriptors of these classes and observe the classification performance of the machine learning algorithm. Second, we design the ground truth model based on the characteristic of each stimulus. We define the distribution of information from the ratio of the face, body, and background area in the stimulus. Lastly, we link them to figure out whether the viewer monkey learned enough about the information in the stimulus.

Keywords: Learning Status Decision, Learning Task Modeling, Looking Behavior Analysis, Looking Pattern Analysis, Behavioral and Social Data Analysis.

1 INTRODUCTION

Like humans, rhesus monkeys live in hierarchically organized societies. Therefore, their life characteristic is interesting to research how their brain-social behavior relationship. To generate an adequate stimulus set, we selected from videos of natural monkeys' behavior segments that could be juxtaposed to mimic dominant-subordinate interaction between pairs of monkeys. In this task, the viewer monkey learns the rank of individuals by observing dominant-subordinate interactions. And we expect there is a specific pattern whether the viewer monkey can extract the rank of individuals from watching videos of pairwise aggressive-appeasing interactions. To measure and quantify the knowledge of social hierarchy, we used the looking pattern (scan paths) of the viewer monkey. Looking behavior (pattern) reflects many things from one's mind (Emery 2000). Not only for human, but also rhesus monkeys are looking where they are interested in even if they have different biological characteristics such as visual acuity. There are several literatures that used looking patterns as a measure of behavior. The monkeys look longer at high-ranking individuals (Deaner et al. 2005) and monkeys show different looking patterns on the images for familiar and unfamiliar individuals (Leonard et al. 2012). Furthermore, monkeys follow the direction of gaze of another monkey individual, thus we can observe the aspects of "The theory of mind" from monkeys (Mosher et al. 2010, Emery et al. 1997, Ferrari et al. 2000, Tomasello et al. 1998). In section 2, we explain the stimulus model and experimental steps which are designed based on these ideas.

One of the important parts (but also the hardest one) when we are doing the behavioral experiment is measuring the viewer monkey's learning status about the task. To decide learning status, we need to quantify the information in the stimuli and figure out how much information the learner has received. Thus, we suggest the way to measure the amount of information in the stimulus and how to decide learning status of the rhesus monkey in the tasks. The main ideas are to determine the appropriate data descriptor and apply it to the machine learning algorithm to classify each type of data sets. Determining the right data descriptor is very important to get a better performance from the machine learning classifier (Lee et al. 2015, Gao et al. 2019). In section 3, we explain how to determine the descriptor model and decide the ground truth data. In addition, we show the results of how we decide the learning status of the viewer monkey about the given tasks.

2 EXPERIMENTAL DESIGN

2.1 Stimulus Sets

We used the artificial hierarchy groups to introduce the hierarchy to the viewer monkey. Each hierarchy group is composed of four different individuals which have the same gender. These individuals are chosen from the movie pool of monkeys in the California National Primate Research Center of UC Davis and the viewer monkeys have never met or seen these individuals before. As shown in figure 1, members in each group are labeled in the descending hierarchy order of Alpha, Beta, Gamma, and Omega. We arbitrarily positioned the most dominant individual as Alpha and the most subordinate one as Omega. In this paper, we introduce the result with the male hierarchy group, which is named Patriline 1 (P1).

To introduce the individuals in the group and to teach their hierarchy positions, we produced the artificial interaction stimuli, which are constructed by presenting the pair of individuals side by side in the 15-second movie. There are six combinations (Alpha-Beta, Alpha-Gamma, Alpha-Omega, Beta-Gamma, Beta-Omega, and Gamma-Omega) and each combination has the flipped version of right and left side to keep the stimulus right-left balanced. In each pair, the individual who has a higher social status, the dominant, threatens the subordinate individual and the subordinate individual shows the appeasing motion to the opponent. As a result, Alpha always threatens another individual in the movie and Omega always makes appeasing expressions to the other side. Beta and Gamma who are located in the middle hierarchy position have threatening or appeasing expressions depending on who the opponent is. Each side of individual scene has 640x480 VGA resolution and the movies are produced with the 1920x1080 full-HD frame size, 25 frames per second of frame speed.

2.2 Experimental Steps

As shown in figure 2, each trial starts when the viewer monkey fixates on the start cue which is shown in the center of the monitor with the sufficient amount of time. Right after the viewer monkey succeeds to fixate on the start cue, the hierarchy movie will play for fifteen seconds. The viewer monkey can look at any place inside or outside of the monitor while the hierarchy movie is playing and the viewer monkey will get the rewards after watching the movie. The amount of these rewards is the same for every trial regardless of the movie type. One block is comprised of twelve trials so that we can show all twelve combinations, six combinations from four monkeys and flipped version of each, in one block. We repeated three blocks in a row and we call it as a session. In the data analysis step in chapter 3, we are going to use the session as a unit of data set. While the viewer monkey is watching the hierarchy movie, we are tracking the viewer monkey's scan paths. Using these scan paths we can measure the looking time, specifically how much time eye fixation stayed on each side of individual in each movie frame and the looking pattern. We will suggest the way to analyze looking pattern in this paper measuring the amount of information that the viewer monkey learned from the hierarchy movie.

3 METHOD

We measured the looking pattern in the scan paths data from the one side of the area of individual pair in the movie so there are six classes; Alpha_TH, Beta_TH, Beta_AP, Gamma_TH, Gamma_AP, and Omega_AP. TH denotes 'threatening' and AP denotes 'appeasing' in the class name. Each class' frame area is 640x480 pixels in the movie and we recorded scan paths with 1kHz rate.

3.1 Counting Fixation Points and Noise Filtering

First of all, we map and count the eye fixation points on the area where the individual monkey showed up by using scan paths data that were recorded while the viewer monkey was watching the movie for fifteen seconds. In equation (1), $E_{frame}(x, y)$ denotes the matrix of the counting of the eye fixation for each frame. We can get a fixation frequency $F(x, y)$ of each class from each trial.

As shown in Figure 3, the window slides entire plane on $F(x, y)$ and it can reduce the measurement noise of eye tracker. The window size is $(2N + 1) \times (2N + 1)$ and we set $N = 20$. This is due to the fact that we set the minimum size that monkey can fixate on the monitor as 1.5 degrees of visual angle. Besides, 40x40 pixel size in the monitor is 1.5 degree of the visual angle in our experimental setup. We set the distance from the monitor to the viewer monkey to 59cm because the size of object in the monitor and visual angle are proportional at this distance (McCready 1985). The filtering window considered all values inside it and output the normalize-filtered value for each pixel as shown in equation (2).

$$F(x, y) = \sum_{frame} E_{frame}(x, y) \tag{1}$$

$$F_{Filtered}(x, y) = \frac{1}{(2N+1)^2} \sum_{i=x-N}^{x+N} \sum_{j=y-N}^{y+N} F(i, j) \tag{2}$$

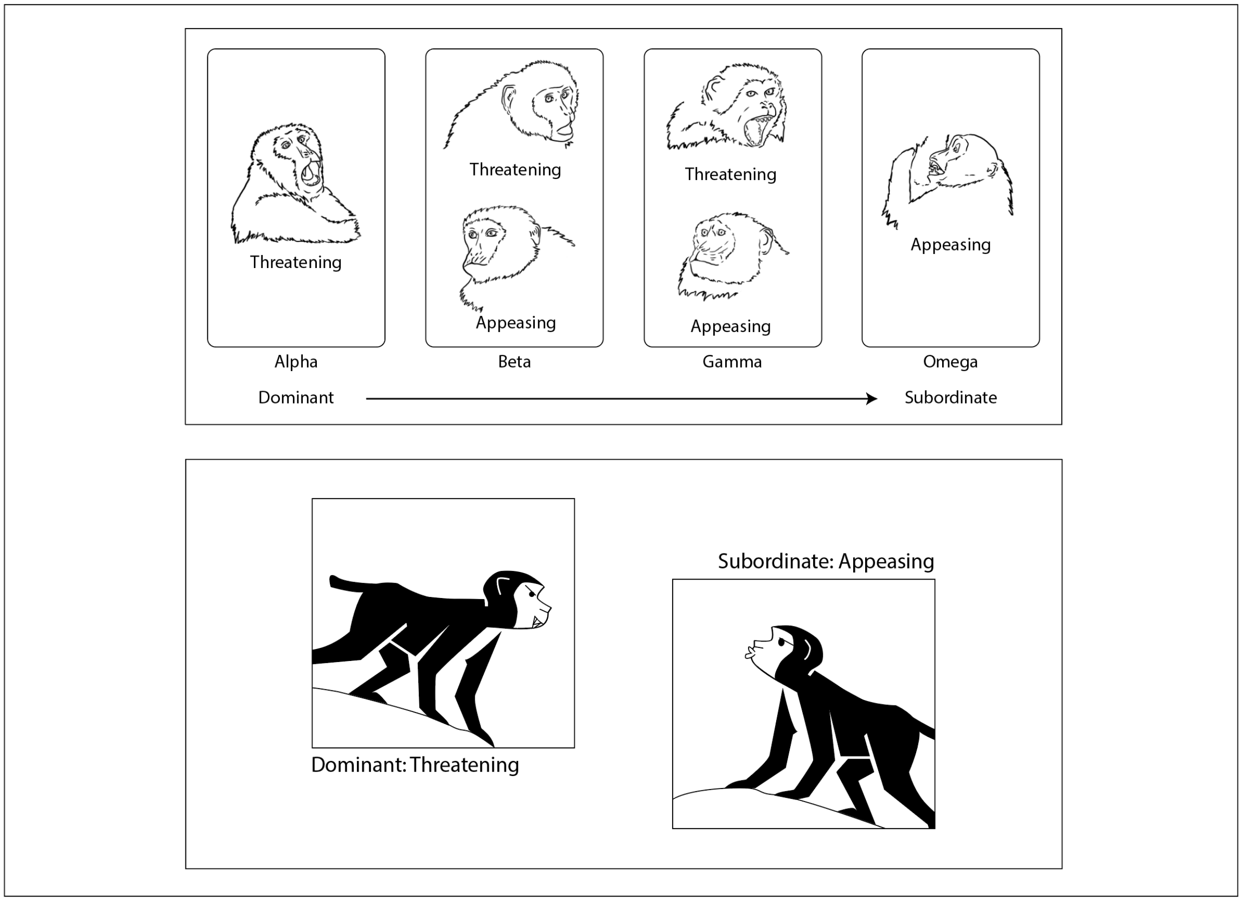

Figure 1: Hierarchy order in the hierarchy group (top) and the illustration of the artificial interaction movie (bottom).

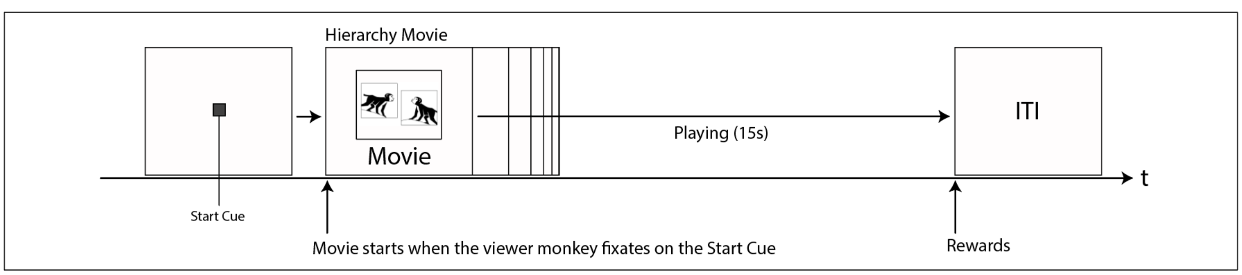

Figure 2: Experimental Steps

3.2 Define of the Descriptor

From the filtered data $F_{Filtered}(x, y)$, we can define the descriptor value with the fixation population on x-axis projection (y-stack) and y-axis projection (x-stack) values. In equation (3), the sequence of descriptor values D is calculated with the label of one of the classes. D is calculated by each trial and each trial produces two descriptor values for two classes in the movie as shown in figure 4. These data sequences of descriptor values will be used for Support Vector Machine (SVM) training.

$$D = \left[\sum_{y=1}^{height} F_{Filtered}(x, y) \;\; \sum_{x=1}^{width} F_{Filtered}(x, y)^T \right] \tag{3}$$

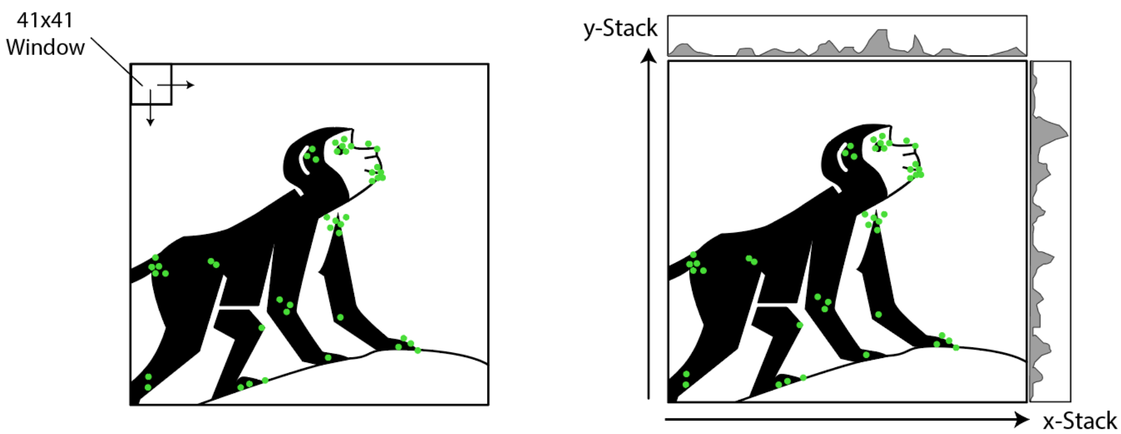

Figure 3: Window slides to denoise and count the fixation points (left), and the algorithm projects counted values on x axis and y axis to make the sequence of descriptor (right). Green dots are the examples of the eye fixation points.

Data_session 1 = {("Alpha_TH", D_{trial1}), ("Beta_AP", D_{trial1}), ... , ("Gamma_TH", $D_{trial\,36}$), ("Omega_AP", $D_{trial\,36}$)}

Data_session 2 = {("Alpha_TH", D_{trial1}), ("Beta_AP", D_{trial1}), ... , ("Gamma_TH", $D_{trial\,36}$), ("Omega_AP", $D_{trial\,36}$)}

⋮

Data_session N = {("Alpha_TH", D_{trial1}), ("Beta_AP", D_{trial1}), ... , ("Gamma_TH", $D_{trial\,36}$), ("Omega_AP", $D_{trial\,36}$)}

Class Labels | Descriptors

Figure 4: The composition of descriptor sequences in each session.

3.3 Ground Truth Data

We defined the amount of information as a ground truth. To measure the amount of information in each frame of the movie, we assumed that the face and body area have more important information than the background area in the frame because these areas contain the appearance characteristic of the individual monkey. This is due to the fact that the probability of fixation is higher in the facial area in the image. (Leonard et al. 2012) To inform the hierarchical structure of the group, the important factor is the face expression. The dominant threatens the subordinate and the subordinate shows the appeasing facial expression in the movie, thus we can expect that the facial area can contain more information than other areas. In equation (4), we define the pixel weight for face, body and background areas as W_{Face}, W_{Body}, and W_{Bckg}. These values are determined by the reciprocal ratio of each area ($\frac{1}{P_{Face}}, \frac{1}{P_{Body}}, and\ \frac{1}{P_{Bckg}}$) in every frame.

As shown in equation (5), we can define the distribution of information $I(x, y)$. If the pixel on the source frame $Src(x, y)$ is the element of face area like as in figure 5-a and 5-b, we will apply the weight value W_{Face} on this pixel. The weight values are applied to other pixels, which are the elements of the body and background area, in the same way.

$$W_{Face} : W_{Body} : W_{Bckg} = \frac{1}{P_{Face}} : \frac{1}{P_{Body}} : \frac{1}{P_{Bckg}} \tag{4}$$

$$I(x,y) = \begin{cases} W_{Face} & (if\ Src(x,y) \in Face) \\ W_{Body} & (if\ Src(x,y) \in Body) \\ W_{Bckg} & (if\ Src(x,y) \in Background) \end{cases} \tag{5}$$

We can understand $I(x,y)$ as $F_{Filtered}(x,y)$ in equation (3) and get a ground truth descriptor value G via equation (6). The ground truth data of each class G is shown in figure 5-c. G is the sequence of the values that are composed to the x-axis projection of column values of $I_{frame}(x,y)$ and transposed y-axis projection row values of $I_{frame}(x,y)$. Lastly, we accumulated these sequences from each frame to one sequence G as shown in equation (6). Another important meaning of G is that it can express the characteristic of stimulus. G will have a wider shape if the object (the individual monkey) in the movie frame moves a lot, and plot's offset value will be higher if the object's facial area is bigger than others. According to these results, we can quantify the characteristic of each stimulus.

To use G as a ground truth value, we need to add simulation modeling on this to compare the classification result with the result from the real data. In equation (7), LT_{trial} denotes the looking time rate on one class (side of the individual) out of fifteen seconds in each trial. LT_{trial} changes the scale of value G every trial and we used the Additive White Gaussian Noise as a noise model for this signal. G will be calculated by each trial and each class so simulated descriptors $\widehat{D}$ will also provide by trial and class as shown in figure 6. We will explain how the simulated descriptors $\widehat{D}$ apply to the classifier and output in the section 3.4.

$$G = \sum_{frame}[\ \sum_{y=1}^{height} I_{frame}(x,y)\ \ \sum_{x=1}^{width} I_{frame}(x,y)^{T}\ ] \tag{6}$$

$$\widehat{D} = LT_{trial}G + AWGN \tag{7}$$

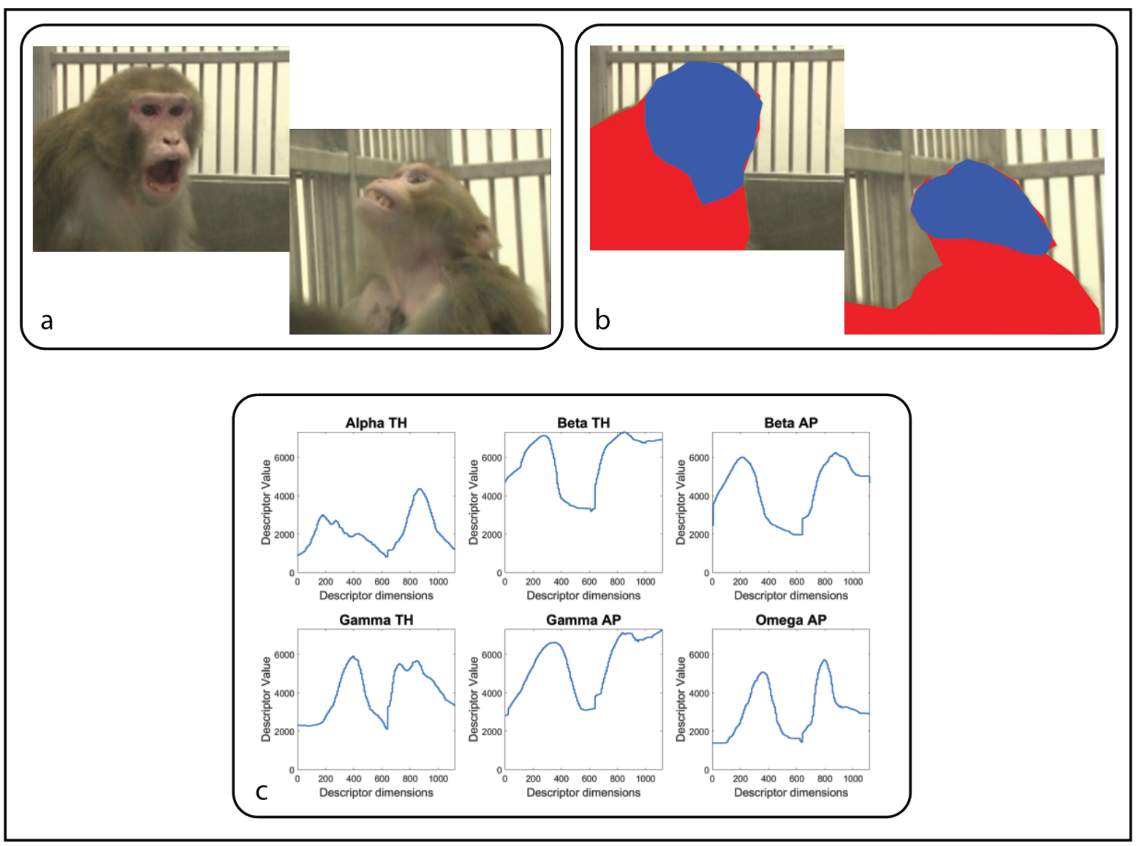

Figure 5: (a) The original frame from the hierarchy movie, (b) Segmented face and body area from the original frame, (c) The ground truth data for each class.

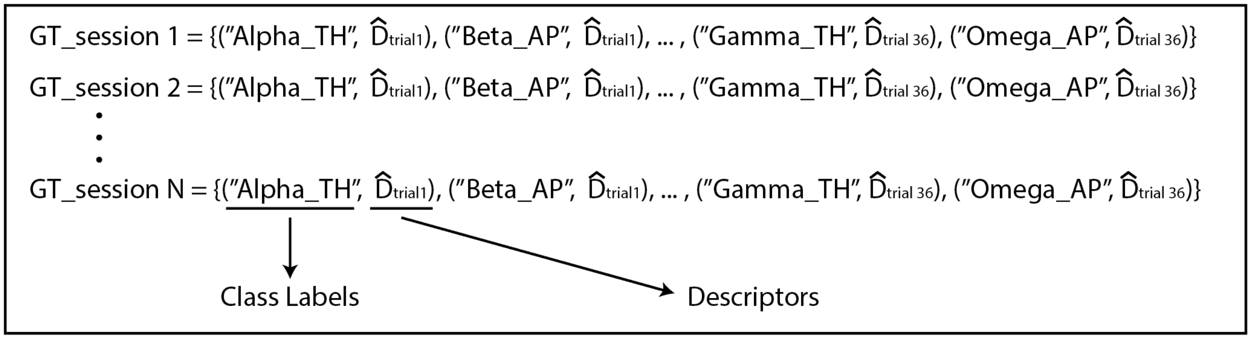

Figure 6: The composition of ground truth descriptor sequences in each session.

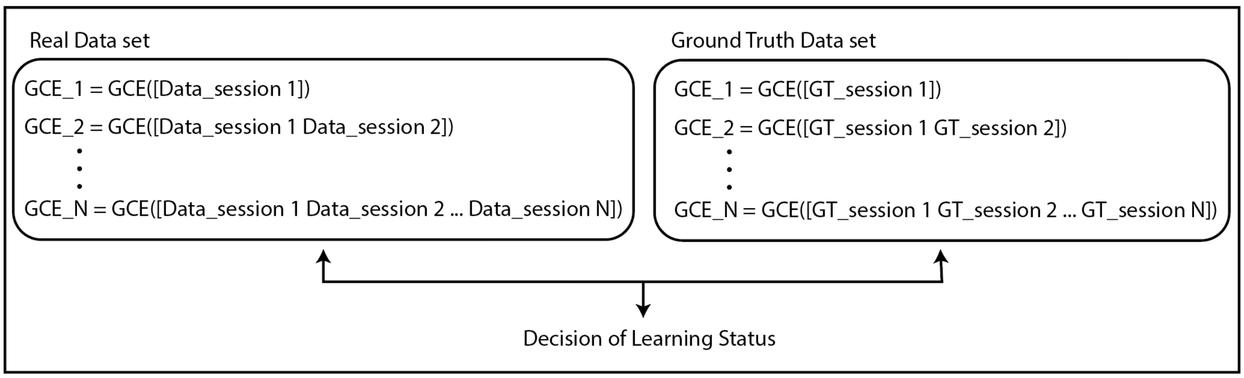

Figure 7: The Comparison of GCE to Decide Learning Status.

3.4 Estimating the Generalized Classification Error

We apply these descriptor values and labels from equations (3) and (7) to SVM classifier. Using the cross-validation scheme, we can estimate the Generalized Classification Error (GCE) and judge how much the data sets from each class are significant. Furthermore, we can decide the learning status of the viewer monkey about the task through the comparison of classification result between real data and ground truth. As shown in figure 7, GCE values will be compared with the GCE values from the ground truth data sets. Sequentially accumulated data sets and their results will show the time point when the classification performance is saturated or overfitted. We are going to use this point to decide the learning status of the task whether the viewer monkey learned this task completely or not. This means the repeated sessions will make the GCE values to follow the patterns from ground truth data sets.

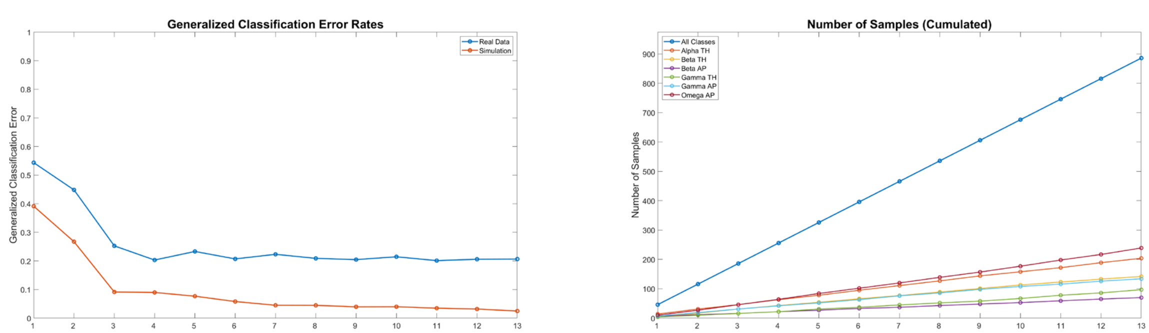

Figure 8: Comparison of GCE rate (left) and the number of samples by accumulated sessions (right).

3.5 Experimental Results

As shown in figure 8, GCE rate shows decreased and saturated tendency in both real data and ground truth data. Classification performance is much better in ground truth data but the trend of the data proceeds very similarly. This is because we can observe that the looking pattern from the viewer monkey is following the pattern of ground truth's in figure S2. Classification performance is saturated after accumulating four sessions not only in real data but also in ground truth data even if we accumulated thirteen sessions with over 800 samples. Therefore, we can decide the learning status of the viewer monkey on this task when the viewer monkey finished four or five sessions.

Absolute GCE values are higher in the real data. This is due to the fact that the data from the real experiment has more distracting factors. Watching and sympathizing with social hierarchy interaction is not a simple process. In the hierarchy movie, there are two individuals interacting and the viewer monkey cannot equally focus on these monkeys. The viewer monkey will focus more on the side that he has more interest in in terms of socially or personally, and the other side is comparably distracted by it. We found this effect from the signals in figure S2. It can affect to the sample data and it occurred the lower performance classify among the subordinate classes Beta_AP, Gamma_AP, and Omega_AP as we are showing in confusion matrices in figure S1. However, there is not this effect in ground truth data, therefore we cannot observe the reduction of the performance at the confusion matrices that are located on the right column in figure S2.

4 CONCLUSION

We suggested the way to figure out how to define the characteristic of each stimulus class. Using this ground truth and formal mode, we can evaluate how much information the viewer monkey learned and we can decide the learning status with this model. We balanced the stimulus in terms of sides. We showed for the same amount of times a regular movie and its flipped version. Also, we showed the movie for the same time that includes all four monkeys in the hierarchy group. Simulation results demonstrate an appropriate performance (under 10% of GCE rate) when we stack more than three sessions as shown in figure 8. It means that the ground truth modeling and its simulation data sets are reflected significant characteristics of each of six classes.

To expand these results, we will repeat the same experiment with different rhesus monkeys, who have different gender, age, and social position to have data with diversification of individuals. To have the diversification of data sets, we have more hierarchy groups which are composed of female groups and male groups. We call the female hierarchy groups "Matrilines" and the male hierarchy groups "Patrilines". We are also planning to set the control experiments that contain non-social objects and neutral expressions.

In the future, we will design several stages of the task with this learning status decision. We will use looking time differences of each class as a looking behavior analysis of the rhesus monkey. Using these results, we can analyze the theory of mind in the viewer monkey's mind and also we can quantify how much differences among the different social positions in the rhesus monkey society. Furthermore, we will analyze scan paths from the viewer monkey to quantify the amount of gaze following from each side of individual in the stimulus. The results of gaze following can be highly related to the theory of mind. Lastly, we will link these behavioral results to neurophysiology. Brain data will allow us to identify the critical areas of the primate brain involved in the social interaction and the production of expressions.

ACKNOWLEDGMENTS

This research is supported by NIH grant 2P50MH100023-06. Thanks to Natalia Jacobson who provides the illustrations of monkey individual portraits in figure 1.

A APPENDIX: SUPPLEMENTAL FIGURES

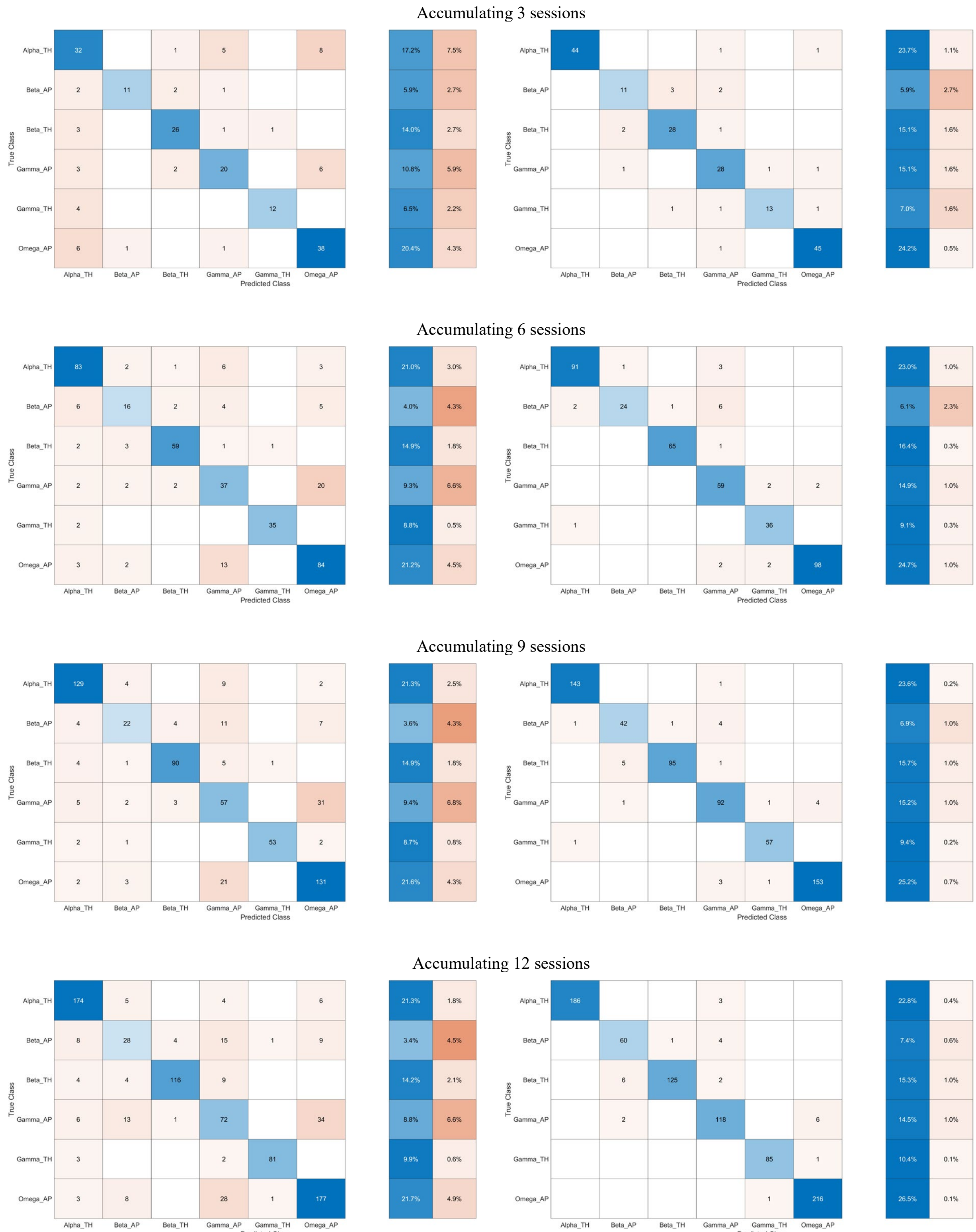

Figure S1: Confusion Matrix with Cross-validation Classification in real data (left), and in ground truth data (right). The percentage tables in each confusion matrix are the classification performances.

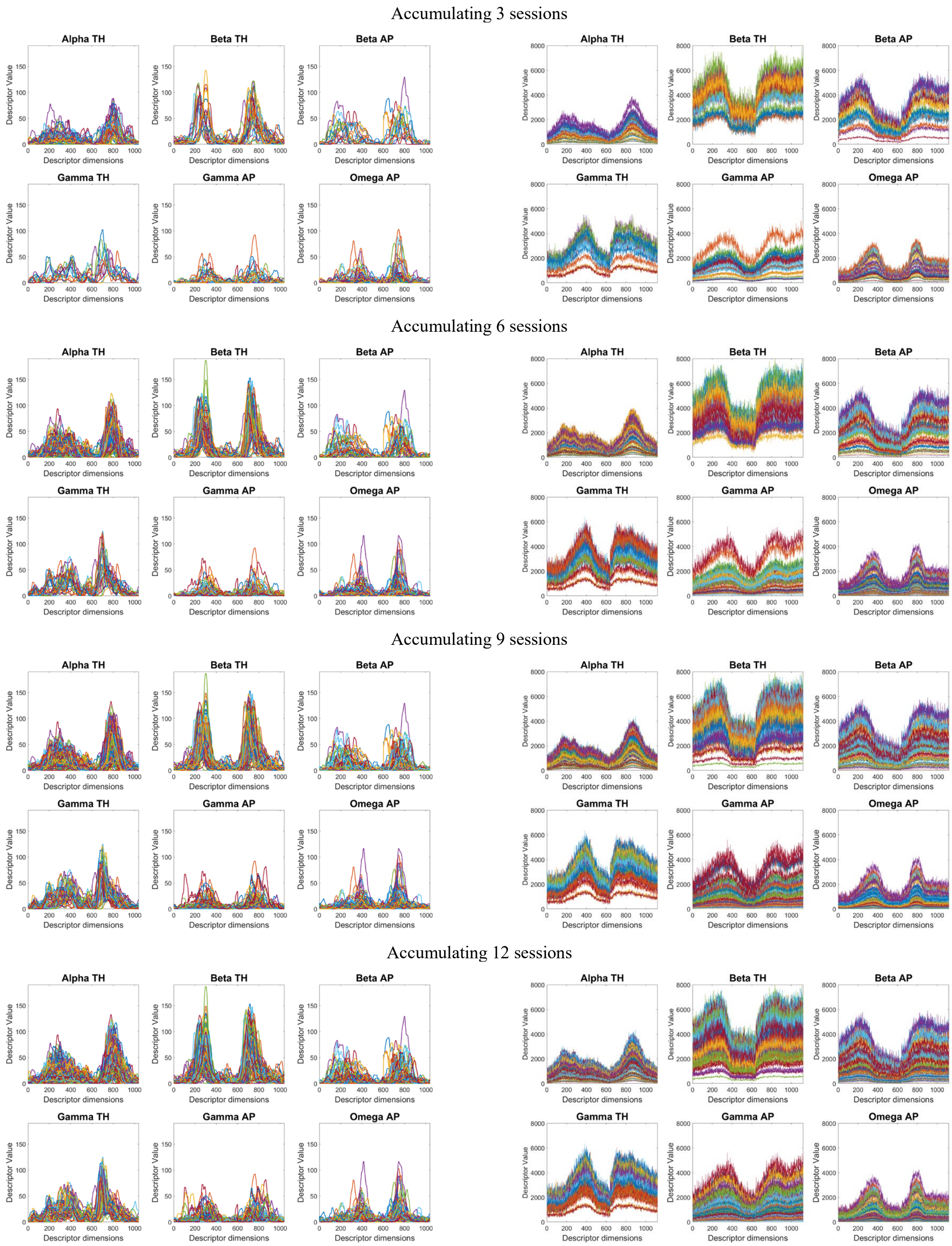

Figure S2: Classification Sample Space of each class in real data (left), and in ground truth data (right).

REFERENCES

Deaner, R. O., and M. L. Platt. 2003. "Reflexive social attention in monkeys and humans". *Current Biology*. Vol. 13. pp. 1609–1613.

Emery, N. J., E. N. Lorincz, D. I. Perrett, M. W. Oram, and C. I. Baker. 1997. "Gaze following and joint attention in rhesus monkeys (Macaca mulatta)". *Journal of Comparative Psychology*. Vol. 111. pp. 286–293.

Emery, N. J. 2000. "The eyes have it: the neuroethology, function and evolution of social gaze". *Neuroscience and Biobehavioral Reviews*. Vol. 24. pp. 581–604.

Ferrari, P. F., E. Kohler, L. Fogassi, and V. Gallese. 2000. "The ability to follow eye gaze and its emergence during development in macaque monkeys". *Proceedings of the National Academy of Sciences, of the United States of America*, Vol. 97, pp. 13997–14002.

Gao, H., J. Wang and J. Sun. 2019. "Improve the performance of machine-learning potentials by optimizing descriptors". *The Journal of Chemical Physics*. Vol. 150.

Lee, S., M. Bang, K. Jung, and K. Yi. 2015. "An efficient selection of HOG feature for SVM classification of vehicle". *2015 International Symposium on Consumer Electronics (ISCE)*. pp. 1-2.

Leonard, T. K., G. Blumenthal, K. M. Gothard, and K. L. Hoffman. 2012. "How Macaques View Familiarity and Gaze in Conspecific Faces". *Behavioral Neuroscience*. Vol. 126. No. 6. pp. 781-791.

McCready, D. 1985. "One size, distance, and visual angle perception". *Perception & Psychophysics*. Vol. 37. No. 4. pp. 323-334.

Mosher, C. P., P. E. Zimmerman, and K. M. Gothard. 2011. "Videos of conspecifics elicit interactive looking patterns and facial expressions in monkeys". *Behavioral Neuroscience*. Vol. 125. pp. 639 – 652.

Shepherd, S. V., R. O. Deaner, and M. L. Platt. 2006. "Social status gates social attention in monkeys". *Current Biology*. Vol. 16. pp. R119–R120.

Tomasello, M., J. Call, and B. Hare. 1998. "Five primate species follow the visual gaze of conspecifics". *Animal Behaviour*. Vol. 55. pp. 1063–1069.

AUTHOR BIOGRAPHIES

SEUNGHYUN LEE is a Ph.D. student who is majoring in Electrical & Computer Engineering and minoring in Neuroscience at the University of Arizona. He earned his B.S. in Electrical Engineering and M.S. in Electronics Engineering from Kookmin University in Seoul, South Korea. His research is covering interdisciplinary perspectives such as analysis of social behavior & neural data and system modeling with computer vision & machine learning algorithm. His email address is eacham@email.arizona.edu.

KATALIN M. GOTHARD is faculty in the departments of Neurology and Neuroscience at the University of Arizona and in the Center for Translational Social Neuroscience at Emory University. She obtained her M.D. in Romania followed by postgraduate training in neurosurgery. In 1996 she obtained her Ph.D. in Neuroscience in the laboratory of Bruce L. McNaughton, and continued her training with postdoctoral work at UC Davis. In 2002 she established her lab at the College of Medicine at the University of Arizona. Her lab has pioneered monitoring cellular brain activity and autonomic responses in the context of natural social behaviors. Her email address is kgothard@email.arizona.edu.

JERZY W. ROZENBLIT is University Distinguished Professor, Raymond J. Oglethorpe Endowed Chair in the Electrical and Computer Engineering Department, and Professor of Surgery in the College of Medicine at the University of Arizona. He holds the Ph.D. and M.S. degrees in Computer Science from Wayne State University, Michigan. Currently, he is developing computer guided training methods and systems for minimally invasive surgery. His email address is jerzyr@email.arizona.edu.