Tracking Gaze Patterns in Human Facial Recognition

Ian Burch
Clemson University

Xiaowei Li
Clemson University

Sam McDade
Clemson University

Kyle Swanson
Clemson University

ABSTRACT

The field of eye tracking contains a body of research which indicates that fixation patterns differ when subjects attempt to recognize a face that is familiar to them vs. when attempting to recognize a face that is unfamiliar. Within the aforementioned body of research are studies which measure this effect when study participants attempt to recognize famous vs. non-famous faces. This paper reviews some of the literature on eye-tracking and facial recognition and reports the results of a simple eye-tracking experiment which attempted to replicate the findings of past studies, particularly one which used famous vs. non-famous faces to elicit from study participants the familiar vs. unfamiliar facial recognition fixation patterns found in the literature.

ACM Reference Format:
Ian Burch, Xiaowei Li, Sam McDade, and Kyle Swanson. 2019. Tracking Gaze Patterns in Human Facial Recognition. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION AND BACKGROUND

Experiments seeking to determine the mechanics of human facial recognition go back at least 50 years, and early experiments quickly made it apparent that humans, when recognizing faces they have seen before, do so with distinctly different gaze patterns than they do when recognizing faces which are unfamiliar to them [5]. This effect is so pronounced, that studies have been using eye tracking technology to detect whether a subject recognizes a face without the subject revealing their prior level of familiarity. Millen et al. conducted a 2017 study [7] in which facial recognition was detected even when participants attempted to hide the fact that they were familiar with the face presented to them, simulating a police interrogation.

Millen et al.'s 2019 study [6] observed that the markers of facial recognition are that familiar faces elicit less fixation on the face as a whole and fewer regions of viewing, with longer fixation duration concentrating on smaller, inner regions of the face. This is backed up by studies such as one conducted by Hsiao et al. in 2007 [4], which showed that when a face is familiar, recognition is nearly instant, and is optimized when the subject looks at one of two points in the center of the face. A proposed mechanism explaining this effect uses the neural network model of face recognition, drawing on the work of Gobbini et al. [3]. Just as in neural networks, when compared to familiar faces, unfamiliar faces require more effort when processed by the human brain, as the brain must collect new information. When attempting to learn a novel face, subjects look around the face, collecting the necessary information for future recognition [7], which will quickly become instantaneous and effortless after a few exposures [2].

In their 1999 study, Althoff et al. [1] used famous vs. non-famous faces as prompts for subjects in an eye-tracking experiment involving facial recognition. Althoff et al.'s study attempted to observe the effects of repeatedly exposing the same faces to subjects, changing them from unfamiliar to familiar, with famous faces as a control. This study does not test repeated exposure. This study merely aims to replicate the more general finding that familiar faces are observed differently than unfamiliar faces, using Althoff et al. as the inspiration for celebrity photos as stimuli.

2 METHODOLOGY

In performing this experiment, our goal was to replicate the findings of past eye-tracking experiments which observed distinct fixation patterns in human recognition of familiar faces. In the study, we used eye-tracking methodology to identify different gaze patterns and fixation duration when looking at familiar vs. unfamiliar faces, with participants being polled to determine whether or not they recognized said faces, as familiar faces were represented by celebrities' faces that were famous in the generation of the selected participants.

2.1 Objective and Hypothesis

This study performed a within-subject (repeated measures) true experiment to test a hypothesis derived from literature review.The objective of this study is to determine the difference in gazing patterns between viewings of familiar vs. viewings of unfamiliar faces. The hypothesis is that high familiarity relates to shorter viewing times and smaller gazing area, while low familiarity relates to longer viewing times and larger gazing area.

2.2 Research Question and Variables

The research question explored was: How does the degree of familiarity affect viewing time and gaze patterns when recognizing well-known and less well-known celebrity faces?

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2.2.1 Independent Variable. The independent variable manipulated in this study was the degree of pre-existing familiarity the participants had with each celebrity facial image. To create variation, a range of major celebrities to very minor celebrities were used for stimulus images. To categorize each subject's individual familiarity with the celebrity images viewed, we used a three-point scale: if the subject has never seen the person before, the degree of familiarity is in category 1; if the subject has seen the person but does not know their name, the degree of familiarity is in category 2; if the subject has seen the person and knows his/her name, the degree of familiarity is in category 3.

2.2.2 Dependent Variable. The dependent variable measured was the gaze interaction of each participant with the areas of interest on each celebrity image, measured as fixation location and duration within the areas of interest. Our AOI's (Areas of Interest) were separated into 3 major areas: left eye, right eye, and mouth, with left and right in this case being from the perspective of one viewing a face. Pre-experiment images were processed so that the faces in each image aligned with the AOI's, which means AOIs (eyes and nose) for each celebrity face were in the same location on the computer screen.

2.3 Subjects

This study featured 11 participants. These participants were volunteers, undergraduate or graduate students, from Clemson University. The participants were screened on the basis of corrective lenses, with those that wore corrective lenses not being allowed to participate in the study. This was done in order to avoid interference from the lenses with the Gazepoint eye tracker. Celebrity knowledge is fairly generation specific and therefore the participants were kept between the ages of 18 and 25 and informally screened for pop culture knowledge to keep the culture of each participant the same. This was done in the hope that each participant would recognize around half of the celebrities. There were no other screens that the participants had to pass to be allowed to participate in the study. The study was conducted as outlined by the Collaborative Institutional Training Initiative and approved by Institutional Reviewing Board. No unexpected accidents occurred and needed to be reported.

2.4 Stimuli

The set of stimuli contained 16 images, selected by searching for clear celebrity headshots with the same resolution. Images were then converted to a 800 by 1050 resolution and displayed in the center of the screen (as in Figure 1). Following this, image alignment was tweaked to fit the locations of the eyes, nose, and mouth in the center of the screen, so there was no difference in the location of the AOI's relative to each image. For the first 3 seconds after exposure to each image, fixations were tracked. This gave the gaze path of the eyes as well as the duration and intensity of fixations in each area of interest. Between stimuli a blank screen with a cross at a random location gave the subject a reset (as in Figure 2).

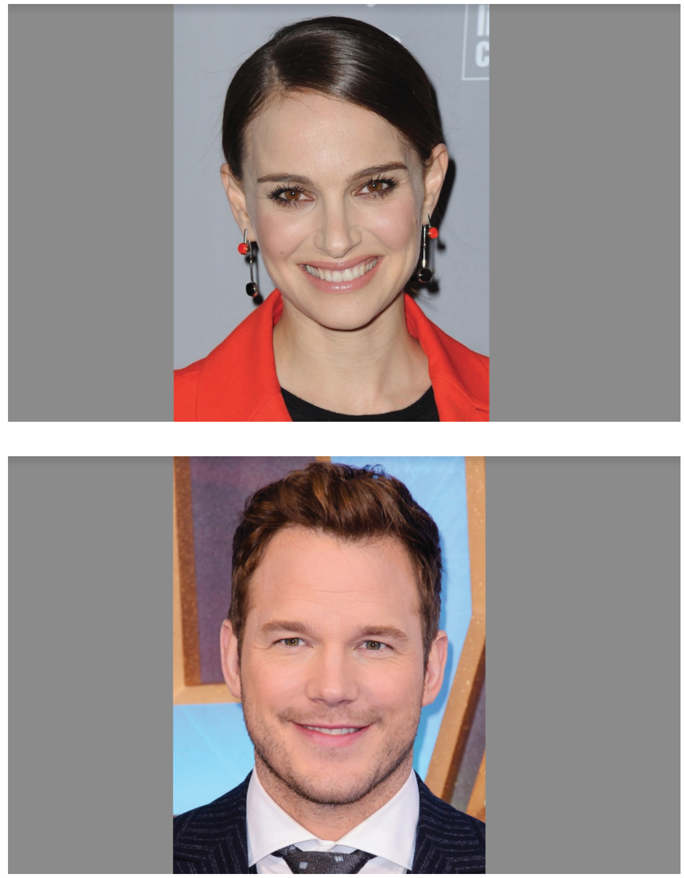

Figure 1: Two of the images used as stimuli.

Figure 2: Blank screen with cross used as a reset.

2.5 Apparatus and Data Collection

Stimuli were displayed to each participant on a 22-inch Dell P2213 LED-backlit LCD monitor with 1680 by1050 pixels resolution, a refresh rate of 60 Hz, and a color depth of 32 bits. Viewing distance was held constant at 21 inches.The equipment used for this experiment to collect data was the pupil corneal reflection video-based Gazepoint eye tracker, GP3. The Gazepoint GP3 is a research-level tracker that has between .5 and 1 degree of accuracy and a 60 Hz sampling rate. This device was used in conjunction with a computer to track users' eyes. The gaze data was recorded using the Gazepoint Analysis software. Prior to conducting the experiment with each participant, the eye tracker was re-calibrated (As in Figure 3). The Gazepoint Analysis software was also used to compile the raw results.

Figure 3: Calibration of the eye tracker.

2.6 Procedure

The study was conducted as a within-subjects (repeated measures) experiment. Participants were shown a sequence of celebrity faces on the display screen while being eye-tracked, with the faces ranging in fame from obscure to very well known. Images were displayed in random order from a selection of celebrity headshots chosen for their uniform nature. Each participant was shown 16 images. Before the presentation of each image, the participant was shown the default screen (As in Figure 2). Then, the image was shown on the screen and the participant was asked to rate the face seen in the image on a scale from 1 to 3 as a researcher recorded their responses, with a response of 3 indicating a face they recognized and could name, 2 a face they recognized but could not name, and 1 a face they did not recognize. Finally, the screen went back to the default screen to prepare for the presentation of the next image. This procedure allowed the experimenters to determine which faces subjects recognized, as well as the strength of that recognition, and record this data.

3 RESULTS

Our collected data was recorded and exported from Gazepoint Analysis. We ran a script to extract specific metric data and conducted statistical significance tests between those metric data and the familiarity scale. We extracted three indicators; transition entropy, fixation duration of each AOI, and the percentage of time spent in each AOI. For the transition entropy, the results were $F(2,20) = 2.324$, $P = 0.124$, indicating no significance. For the duration spent on each AOI the results were $F(4,14) = 6.094$, $P = 0.1459$, indicating no significance. For the percentage of time spent on each AOI the results were $F(2,20) = 0.28$, $P = 0.759$, indicating no significance.

To analyze the difference in the observed viewing patterns, we examined the sum of fixation duration on our three AOIs for each subjective familiarity rating 1, 2, or 3 across all subjects and stimuli (Figure 4). Difference in fixation duration was analyzed and tested for significant difference (Fixation Duration Analysis), and we found that there is no significant difference in fixation duration on the AOI's when a subject is viewing a familiar face vs. when the subject is viewing an unfamiliar one. The results were $F(4,14) = 6.094, P = 0.1459$.

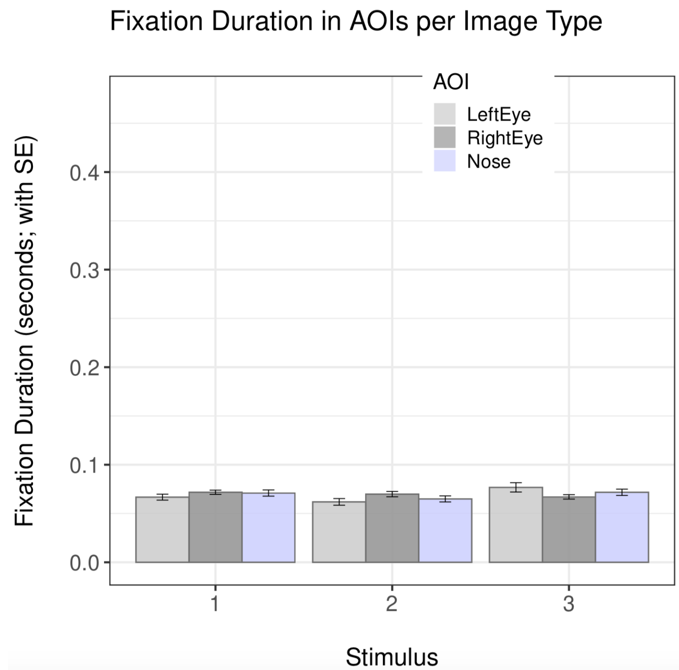

Figure 4: Fixation Duration Analysis

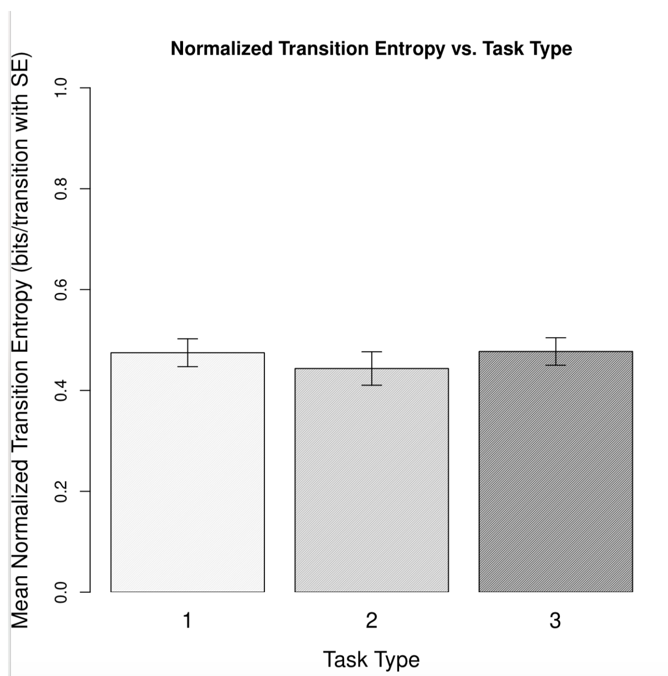

Figure 5: Mean Transition Entropy

Transition entropy is a measure of the path the subject took over the AOI's when viewing a face. It indicates the viewing strategy of the subject. We looked at mean transition entropy across all subjects and stimuli for each subjective familiarity rating of 1, 2, or 3 (Figure 5). For the mean transition entropy the result was F(2,20) = 2.324, P = 0.124, indicating there is no significant difference in subjects' viewing paths when looking at familiar vs. unfamiliar stimuli.

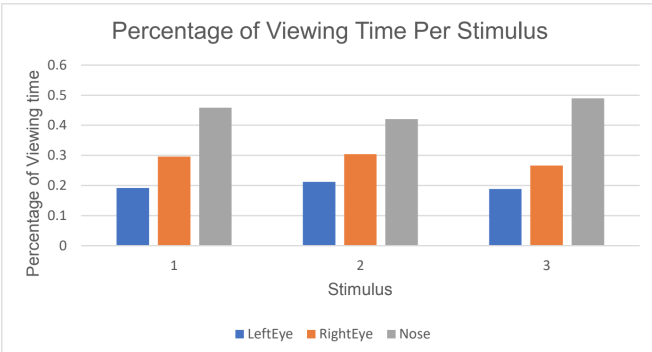

Figure 6: Percentage of Viewing Time per AOI Analysis

The Percentage of Viewing Time per AOI Analysis (Figure 6) shows which AOI subjects were looking at as a percentage of time on average as they were viewing the stimuli broken down by subjective familiarity rating. The results for significance were, F(2,20) = 0.28, P = 0.759, indicating there is not a significant difference in the way subjects looked at the AOI's in images with differing subjective familiarity levels. However, the results do seem to show a pattern, in that subjects tended to look between the right eye (from a viewer's perspective) and the nose when trying to recognize a face.

A further visualization of this result can be seen in Figure 7.

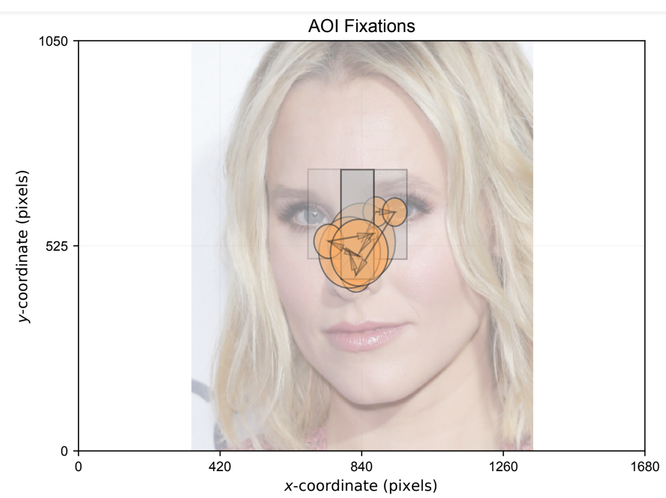

Figure 7: Average AOI Fixation Visualization

4 DISCUSSION AND CONCLUSION

Our hypothesis was that faces which subjects were highly familiar with would record shorter viewing times and smaller gazing area, while faces with low familiarity for the subjects would have longer viewing times and larger gazing area. Through our results, we found that there was no significant difference between familiarity and unfamiliarity with regards to fixation location and duration. Therefore, our hypothesis was found to be incorrect. This study failed to show significant results that back up the literature on the difference in fixation and fixation duration. However, there are still valuable results. Our viewing pattern did match a pattern described in the literature; we found that our subjects tended to recognize faces by fixating on the area between the right eye (left eye of the person, right eye from a viewer's perspective) and nose [4].

There are a few possible issues with our study which might explain why we found no significant difference. One reason could be that we used an insufficient sample size (N = 11). Another reason is that we may not have set a clear evaluation rubric for familiarity. Subjects could be confused about their level of familiarity during the experiment. In any future studies, a valid standard of measuring familiarity is needed. A third possible reason is that we may have focused too much on AOI's in the center of the face. Perhaps AOI's placed elsewhere might have caught differences in fixations outside the center of the face, though similar times spent in each AOI observed in this study imply that the time spent on fixations outside the face was similar in duration, if potentially different in nature, when viewing familiar vs. unfamiliar faces. Further, it is possible that centering the eyes of each face used as a stimulus in the center of the screen was too arresting and drew subjects' eyes towards the center of the face, with this perhaps being compounded as they "learned" to look to the center of the screen after each reset. Another potential issue is that even the minor celebrities were somewhat recognizable, perhaps many of the participants had seen most of the faces before, even if they were not fully conscious of it. Finally, it is possible that subjects being told to focus on determining whether or not they knew the celebrity in the stimulus, and being given a limited amount of time to do so, made them use their three seconds of viewing time to fixate on the most useful part of the face for facial recognition, the center, whether they recognized the celebrity or not. A study with a more passive recognition procedure might find different results.

4.1 Validity

Before we performed the real study, we conducted a pilot study on one subject to get familiar with procedure and revise steps based on this test. This increased the internal validity of study. The experiment has an innate shortage of external validity, since it has not been replicated and contradicts the findings of existing research, but we have released all of our methods, procedures, and results for further analysis and potential replication.

Due to the time and cost, our sample size was limited to 11 subjects. A larger sample size would have given us higher statistical validity. Although we randomized the viewing position by putting the black screen with a cross image in between each stimulus, we did not randomize the order of the stimuli. This is not a significant problem since each subject only underwent the experiment once.

Order randomization should not make a significant difference in this study, but was a potential factor in decreasing validity.

4.2 Final Thoughts

Although our hypothesis was that higher familiarity relates to shorter viewing times and smaller gazing area, while low familiarity relates to longer viewing times and larger gazing area, our study found no significant difference in three outcomes including transition entropy, fixation duration of each AOI, and the percentage of time spent in each AOI. However, this study successfully extracted similar viewing patterns as seen in literature which suggests that subjects tend to look between the right eye (from a viewer's perspective) and the nose when trying to recognize a face. Also, the study set up a replicable procedure for future studies.

REFERENCES

[1] Robert R. Althoff and Neal J. Cohen. 2019. Eye-movement-based memory effect: A reprocessing effect in face perception. 25, 4 (July 2019), 997–1010. https://doi.org/10.1037/0278-7393.25.4.997

[2] Benjamin Balas, David Cox, and Erin Conwell. 2007. The Effect of Real-World Personal Familiarity on the Speed of Face Information Processing. 2, 11 (Nov. 2007), e1223. https://doi.org/10.1371/journal.pone.0001223

[3] M. Ida Gobbini and James V. Haxby. 2007. Neural systems for recognition of familiar faces. 45, 1 (Jan. 2007), 32–41. https://doi.org/10.1016/j.neuropsychologia.2006.04.015

[4] Janet Hui-wen Hsiao and Garrison Cottrell. 2008. Two Fixations Suffice in Face Recognition. 19, 10 (Oct. 2008), 998–1006. https://doi.org/10.1111/j.1467-9280.2008.02191.x

[5] Robert A. Johnston and Andrew J. Edmonds. 2009. Familiar and unfamiliar face recognition: a review. 17, 5 (July 2009), 577–596. https://doi.org/10.1080/09658210902976969

[6] Ailsa E. Millen and Peter J. B. Hancock. 2019. Eye see through you! Eye tracking unmasks concealed face recognition despite countermeasures. 4, 1 (Aug. 2019), 23. https://doi.org/10.1186/s41235-019-0169-0

[7] Ailsa E. Millen, Lorraine Hope, Anne P. Hillstrom, and Aldert Vrij. 2017. Tracking the Truth: The Effect of Face Familiarity on Eye Fixations during Deception. 70, 5 (May 2017), 930–943. https://doi.org/10.1080/17470218.2016.1172093