Gaze Patterns During Emotion Recognition in Animated Point-Light Facial Displays

Ryan Canales
Clemson University
Clemson, SC, USA
rcanale@clemson.edu

Matias Volonte
Clemson University
Clemson, SC, USA
mvolont@g.clemson.edu

Andrew Duchowski
Clemson University
Clemson, SC, USA
duchowski@clemson.edu

Figure 1: Stimuli used in the study. From left to right: neutral, anger, contempt, disgust, fear, happy, sad, surprised.

ABSTRACT

Biological motion is defined as motions made by living organisms, including motions of individual parts, such as fingers, hands, feet, and so on [2]. Point light displays are frequently used to study human perception of biological motion, especially motions from other humans. These displays are typically made by either physically or digitally placing lights or points at key places on the body and making these points the only visual stimulus of the motion being performed. The idea behind point light displays is to isolate the kinematics of these motions by limiting the visual information provided to only the points of light on a black background. In this study, we constructed a virtual point-light facial display and used it to create the following eight emotional expressions: neutral, angry, contempt, disgust, happy, fear, sad, and surprise. We then created ten second animations of the point light face interpolating from the neutral expression to one of the eight expressions. Each clip was shown to every participant ($N = 9$) three times in random order. We collected eye tracker data from each participant while they viewed each clip, the time it took for them to guess each expression, and the accuracy of their guesses. The results were that although overall accuracy was fairly low (< 60%), the median accuracy increased per trial and the amount of time that the participants took to guess the expressions decreased per trial. Additionally, we found that most fixations were around the eyes of the virtual face, followed by the nose, and the mouth.

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

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/0000001.0000001

CCS CONCEPTS

• **Applied computing → Psychology**;

KEYWORDS

Emotional expressions, motion perception, eye tracking

ACM Reference format:
Ryan Canales, Matias Volonte, and Andrew Duchowski. 2018. Gaze Patterns During Emotion Recognition in Animated Point-Light Facial Displays. In *Proceedings of ACM Conference, Washington, DC, USA, July 2017 (Conference'17),* 4 pages.
https://doi.org/0000001.0000001

1 INTRODUCTION

Facial expressions are key to our interpretation of how someone might feel, and they are especially unique in that they can be considered as universal signals for distinguishing basic emotions [3]. Given no other information but the expression, people can infer the emotional state someone else is in, though cultural differences may alter the interpretation for some expressions [5]. Results from previous work using eye tracking with static faces are that people generally look at five distinct regions of the face when trying to identify the emotional expression. These regions are the eyes, upper nose, lower nose, upper lip, and nasion [8]. Not only do we read emotions from facial expressions; we also gather information from the facial movements to aid in comprehension of speech [7]. Although previous studies have provided useful information and insight into emotional expression identification and perception of the face, the question remains of how we use the information from facial movements alone to aid in discerning expressions. This leads to the main motivation for our study, and that is to try to discover how we use facial movements to decode expressions and how well we can do so. To gain insight into the answer to this question, we used animated virtual point light facial displays to isolate the movements of the face and used eye tracking data to analyze which areas of the face contribute to our ability to discern between different expressions. We expect that the eyes will play an important role in discerning the expressions, and therefore there will be more fixations around the eyes than in either the nose or mouth. We also expect, based on work by Anderson et al [1], that the happy expression will be identified the quickest and the most accurately.

2 METHODOLOGY

2.1 Stimuli

The stimuli used in our study was a set of eight, ten second videos of a point light face. The length of the videos is based on the length of the dynamic expressions used by Krejtz et al [6]. In each video, the face either remained with a neutral expression or was animated from the neutral expression to one of seven expressions. The seven other expressions used in this study were anger, contempt, disgust, fear, happy, sad, and surprise. Figure 1 shows the final frame of the videos for each expression. This set of videos was displayed three times to every participant, with the order of the videos in each set randomized.

2.2 Apparatus

The eye tracking device used was the Gazepoint eye tracker, which gathers data at 60hz. The Gazepoint can track where the user is looking on the screen, with a degree of accuracy of half a degree to one degree. The Gazepoint tracks the gaze of both eyes, and the distance of the eyes from the screen. It can tell when the user blinks, or winks, and which eye winked, all without distracting the user.

2.3 Participants

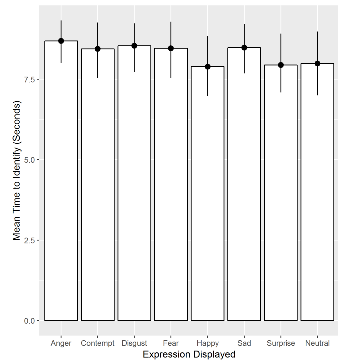
Data was collected from nine college aged (18+) participants.

2.4 Procedure

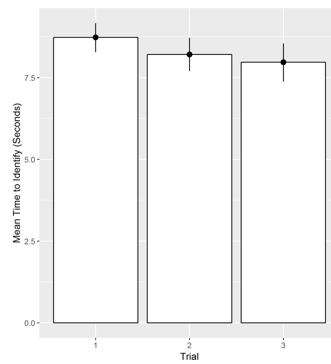
Each session began with the participant performing a standard five point calibration of the Gazepoint eye tracker. Once the eye tracker has been successfully calibrated, the participant could start the playback of the first set of videos by pressing the spacebar. Building upon the procedure used by Gehrer, et al [4], the participants could press the spacebar on the keyboard when they think that they have recognized the final facial expression in each clip. After each clip has been stopped, either by the participant or because the clip has finished, the time taken is recorded, and the participant is prompted to select from a list which expression they thought was in the clip. The order that the expressions are listed in is randomized after each video clip to minimize bias. After the participant has recorded their choice, they can continue onto the next clip by clicking a button on the screen. After the set of videos has been displayed, the eye tracker is recalibrated, and the same procedure is repeated until the set has been viewed three times.

3 RESULTS

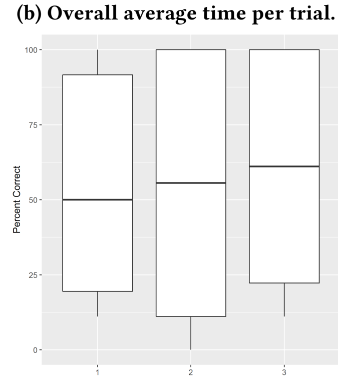
To measure the performance of the participants in reading the expressions, the time taken and percentage of correct guesses were used. Overall average time per expression was computed as well as average time for each expression per trial. Results from a two-way repeated measures ANOVA test showed a significant effect of trial on time $F(2,3) = 3.745, p < 0.05$. This suggests that there is a learning effect; after viewing each expression once, the participants



(a) Mean time for each expression.



(b) Overall average time per trial.



(c) Percent correct per trial.

Figure 2: Results Summary

were generally more quick to guess what the expression shown was. This effect is shown in Figure 2(b).

Another metric measured was accuracy. The expression that was guessed the least accurate was "fear" (it was guessed correctly only 7.4% of the time), with none of the participants identifying it correctly during their second trial. This was due to participants often confusing "fear" with either "surprise" or "neutral". "Surprise" was selected by participants 70.37% of the time that "fear" was shown, and "neutral" was selected 18.52% of the time. Two expressions, "happy" and "surprised", were guessed correctly 100% of the time. "Nuetral" was guessed correctly every time it was shown during the second and third trials, but not the first. This could be due to confusion when first seeing the clip; the video appears to be a still image, unlike the other videos. "Sad" was notably identified correctly 81.5% of the time. Aside from "happy", "neutral", "sad", and "surprise", the expressions displayed in this study were particularly susceptible to confusion This is most likely due to the limited visual information in the point light facial display. "Contempt" was often confused with "happy", with the participants selecting "happy" 62.96% of the time. Interestingly, "anger" was interpreted as "sad" 44.4% of the time and "disgust" was interpreted as "anger" 55.56% of the time. Results from a one-way repeated measures ANOVA did not show any significant effect of trial on accuracy.

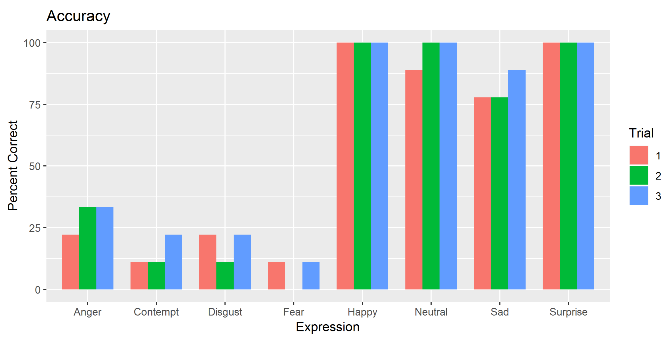


Figure 3: Percent correct for each expression per trial.

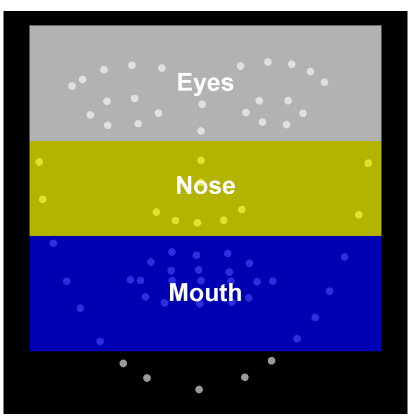


Figure 4: AOI placement on virtual face.

To measure where on the face viewers looked to read each expression, fixations from each participant were computed for each clip, and each fixation was classified as being in one of three areas of interest. To get the fixations, the raw data was first smoothed then differentiated using the Savitzky-Golay differentiation filter. The resulting velocities of the saccades were thresholded such that the corresponding points for velocities below the threshold were classified as fixations. The areas of interest (AOI) defined in this study are represented by rectangles encompassing the area around the eyes, the nose, and the mouth, as shown in Figure 4.

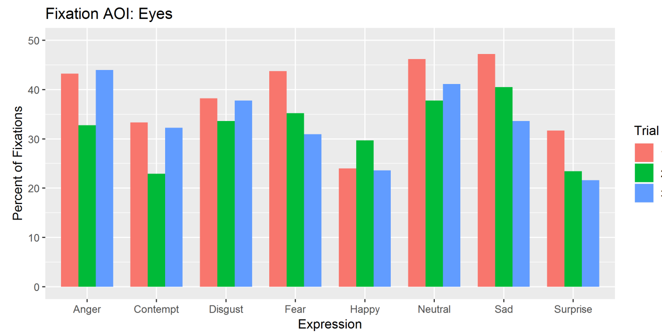


Figure 5: Percentage of fixations around eyes per trial.

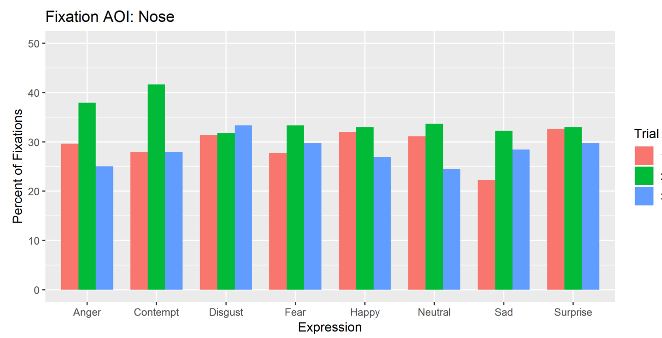


Figure 6: Percentage of fixations around nose per trial.

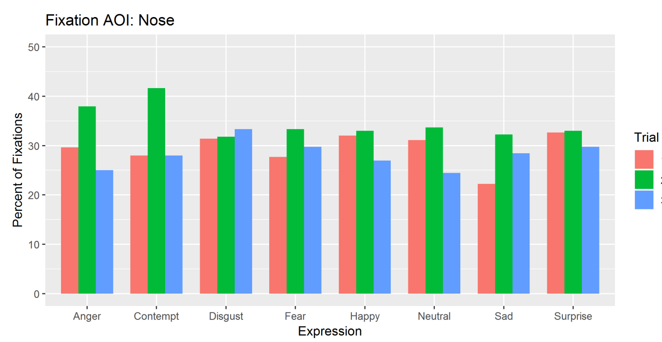


Figure 7: Percentage of fixations around mouth per trial.

The percentage of fixations that fell in at least one of the areas of interest was computed, as were the percentages of fixations within each AOI. Overall, 87.83% of fixations fell in at least one of the AOI's. 35.65% of fixations fell within the AOI around the eyes, 28.52% within the AOI around the nose, and 23.65% within the AOI around the mouth. One interesting effect (though not measured here) of trial on where the fixations landed is that for the expressions "fear", "sad", and "surprise", the percentage of fixations within the eyes AOI decreased for each trial. This can be seen clearly in Figure 5. Looking at Figures 5, 6, and 7, other possible effects of trial on fixation locations could be interpreted. For example, fixations during the expressions "anger" and "contempt" that fall within the

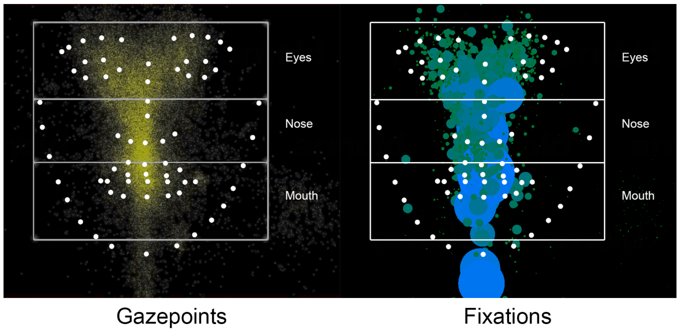


Figure 8: Aggregate gazepoints (sampled points where users looked) and fixations for all expressions.

nose or mouth AOI's increase for the second trial, but then sharply decrease during the third trial. However, further research is needed to determine if these effects are significant and if they hold for a larger number of participants.

Examining where the participants looked is most easily done by looking at visualizations. The left image in Figure 8 shows where on the face all the participants looked over the entirety of the experiment. The right image in Figure 8 shows where on the face fixations (represented by filled circles) occured. The brightness and diameter of the circles representing the fixations correspond to the duration of the fixations. As expected, both the sampled points and the fixations fall mainly on the eyes, the nose, and the mouth. Outliers include the larger circles below the chin, which are most likely due to the eye tracker losing tracking during one of the trials.

4 DISCUSSION

Half of the expressions used in this study were often confused with either sad, happy, or surprised. As hypothesized, happy was identifed very well, as were sad, surprised, and neutral. The fact that there was an effect of trial on time but no significant effect of trial on accuracy suggests that the participants had only improved on identifying the expressions they were able to identify correctly the first time. Where the participants looked on the face followed typical scan patterns around the eyes, nose, and mouth.

4.1 Limitations

This study has several limitations, most notably that it did not include corresponding videos of real faces to compare to the point-light faces. Other limitations include the small number of participants, and the limited statistical analysis. Further research with more participants and with videos of real faces is needed to extract further results, as well as perhaps a higher accuracy eye tracker.

4.2 Conclusions

From this limited study, there are only a couple of concrete conclusions that can be drawn. One is that the expressions "happy", "sad", "neutral", and "surprised" can generally be identified fairly accurately. However, these expressions were also often selected whenever the other expressions were shown. This perhaps indicates that facial movements, at least with a point-light display, are readable, but limited in the range of expresssions easily identifiable. Although this study was very limited, hypotheses for future studies can be drawn from this one. One hypothesis is that people will generally be slower at identifying an expression from a point-light face than from a video of a real face. Another is that although the accuracy will be lower when point-light faces are used, there may be a slight correlation between the best performing expressions in the point-light face and those in a real face. For example, "happy" would yield higher accuracy and shorter times in videos of real faces, as it did with the point-light face in this study. Future work can be done to verify or refute these hypotheses.

REFERENCES

[1] A.K. Anderson, Panitz K. Christoff, and E. De Rosa & J.D. Gabrieli. Neural correlates of the automatic processing of threat facial signals. *Journal of Neuroscience*, 2003.

[2] Randolph Blake and Maggie Shiffrar. Perception of human motion. *Annual Review of Psychology*, 58(1):47–73, 2007. PMID: 16903802.

[3] P. Ekman and W. V. & Friesen. Unmasking the face: A guide to recognizing emotions from facial clues. *APA*, 1975.

[4] Nina A. Gehrer, Andrew T. Duchoswki, and Krzysztof Krejtz. Implementing innovative gaze analytic methods in clinical psychology. *ETRA '18: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, 2018.

[5] Rachael E. Jack, Caroline Blais, Christoph Scheepers, Philippe G. Schyns, and Roberto Caldara. Cultural confusions show that facial expressions are not universal. *Current Biology*, 19(18):1543 – 1548, 2009.

[6] Krzysztof Krejtz, Katarzyna Wisiecka, Izabela Krejtz, PawełHolas, MichałOlszanowski, and Andrew T. Duchowski. Dynamics of emotional facial expression recognition in individuals with social anxiety. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, ETRA '18, pages 43:1–43:9, New York, NY, USA, 2018. ACM.

[7] L.D. Rosenblum and H.M. Johnson, J. A. & Saldaña. Point-light facial displays enhance comprehension of speech in noise. *Journal of Speech and Hearing Research*, 1996.

[8] M. W. Schurgin, J. Nelson, S. Iida, H. Ohira, and S. L. Chiao, J. Y. & Franconeri. Eye movements during emotion recognition in faces. *Journal of Vision*, pages 1 – 16, 2014.