Analysis of Eye Fixations During Emotion Recognition in Faces

REU fellows: Forest Elliott[1]
Faculty Mentors: Dr. H. Cao[2]
Affiliation: [1]Oberlin College, [2] School of Engineering and Computing Science, NYIT
Emails: forest.elliott@oberlin.edu, hcao02@nyit.edu

ABSTRACT

Automatic emotion recognition is a challenging task as emotions can be expressed in a variety of different ways. Most recognition studies have created audio-visual models for classifying emotions. In this study, a new model is proposed based off of human emotion perception. Participants were asked to identify emotions expressed in videos from the CREMA-D database while their gaze patterns were tracked by the Tobii X3-120. The participants watched to sessions of videos, one session having videos in random order and the other having the actors appear in sequential order. The participants gaze patterns were analyzed by looking at their fixation times on specific facial features of the actor in the videos. A decision tree was then used to select the most important facial features needed for a person to recognize negative, positive and neutral emotions. The data mining tool WEKA is used to generate this tree by implementing the J-48 algorithm for generating classifiers. Using this method, classification rates of 63.9%, 64.8%, 66.6% and 69.6% were obtained for sequential, random, correctly perceived sequential, and correctly perceived random respectively.

BACKGROUND & MOTIVATION

Automatic Emotion Recognition

- The ability for computer to automatically recognize human emotion
- Tackled through a variety of audio/visual channels[1]
- Applications in Education, HCI, RHI and Medicine[2]

Action Units (AU)

- Encode Facial Movements
- Combine them to represent emotions [3]

PROBLEM & OBJECTIVE

Research problem:

- Do humans focus on specific facial movements and inflections when perceiving emotions?
- Can this data be used to create an accurate model to predict emotions being perceived?

Objective:

To determine if humans perceive emotions by focusing on specific AU, and build a model that predicts the stimulus emotion based off these fixations

EXPERIMENTAL SET-UP

Record Eye Tracking and Fixation data with:

Video Stimulus of Actor Portraying Emotion

Tobii Eye Tracker

Questionnaire Asking Perceived Emotion

As Participants Watch Videos

- Eyes Tracked
- Are asked what emotion they perceived
- AOI Fixation recorded
- Two sessions,: Random and Sequential

PARTICIPANT RESULTS

Fig. 1: Confusion Matrices of Participants

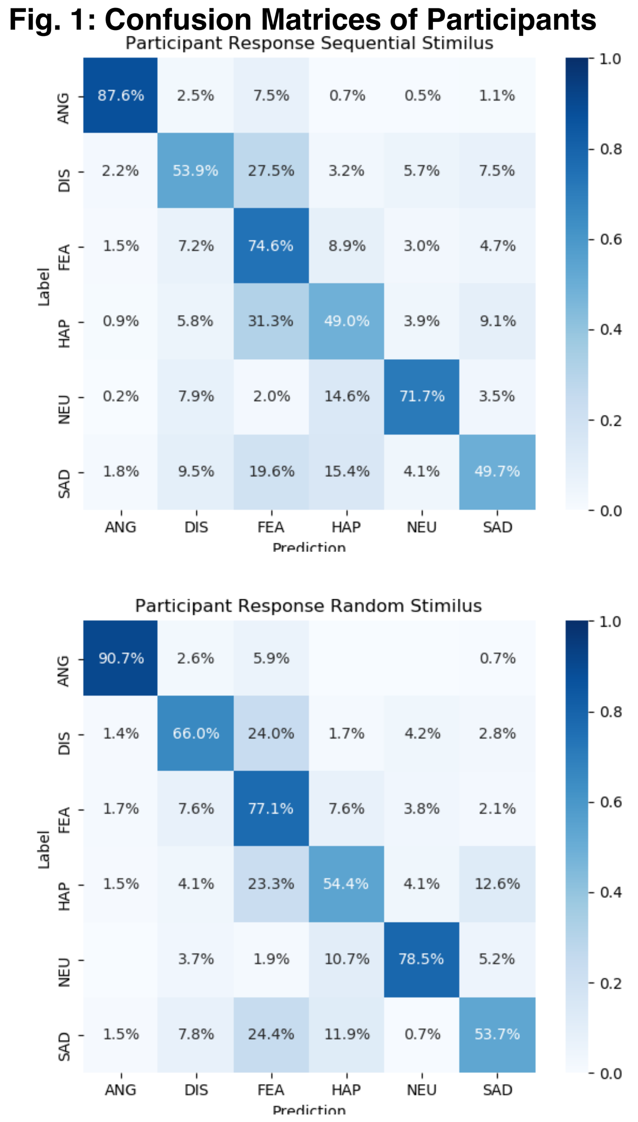


Fig. 2: Areas of Interest on Video Stimulus

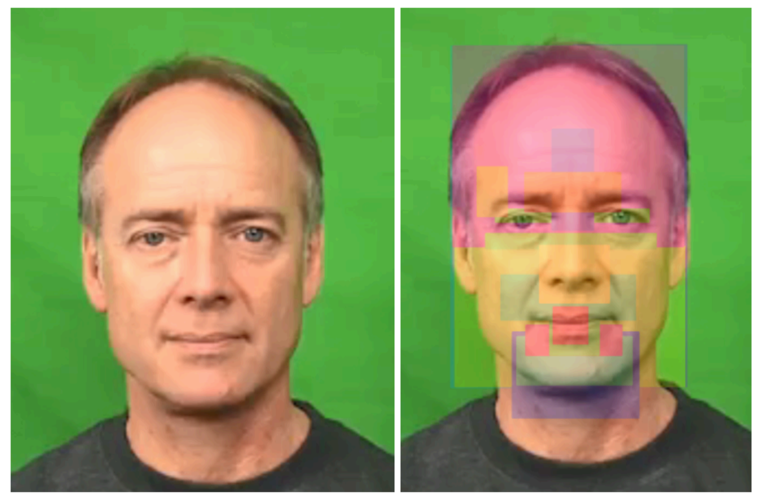


Table I: Average Perception Rate of Participants for Random and Sequential Sessions

Stimulus Order	Average Perception Rate
Random	58.1%
Sequential	70.1%

Table II: Results of Statistical Tests: AOI with Statistically Different Means Fixation Time for Emotions (Negative, Positive, Neutral)

Stimulus Order	AU with Significantly Different Mean Fixation Time
Random	AU1, AU4, AU6, AU7, AU9, AU12, Left face, Right face, Lower face, and Upper face
Sequential	AU7, AU9, Left face, Right face, and Upper face

MACHINE LEARNING & RESULTS

Decision Trees Used for Machine Learning

- Data Pruning and Attribute Selection
- J-48 Algorithm in WEKA [4]
- Divided data into entire data set and only correctly perceived data
- 10-fold cross validation for evaluating model

SMOTE (Synthetic Minority Over-sampling Technique)

- Imbalance of data classes (Negative, Positive, Neutral)
- SMOTE makes new synthetic points to balance classes for machine learning

Full Data Set:

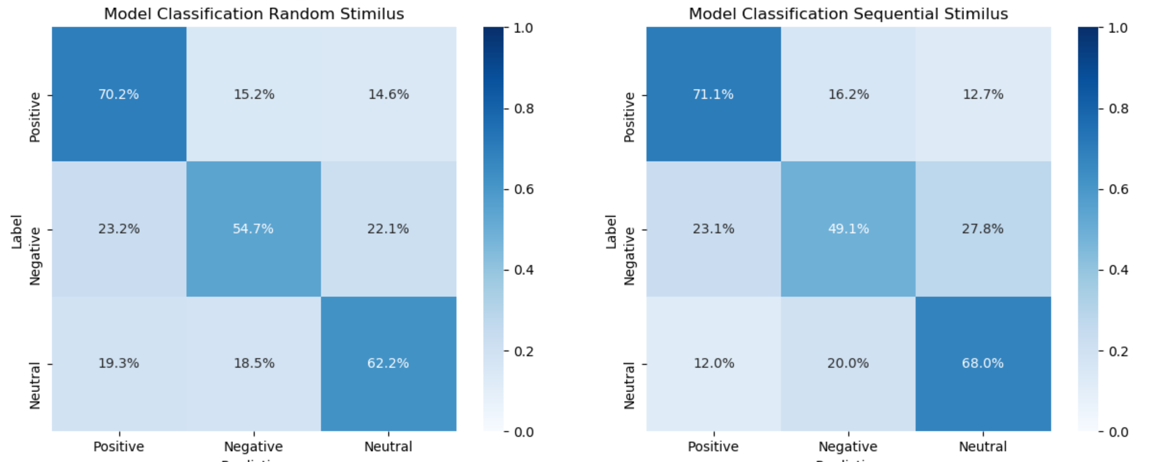
Table III: Classification Rate of Decision Tree on Entire Data Set

Stimulus Order	Classification Rate of Decision Tree
Random (All data)	64.8%
Sequential (All data)	63.9%

Table IV: Selected AOI for Machine Learning on Entire Dataset

Stimulus Order	Selected AOI
Random (All data)	AU6,AU7,AU9,AU12,AU23,AU26, Left,Lower,Right,Upper
Sequential (All data)	AU1,AU2,AU4,AU6,AU7,AU9,AU12 AU15,AU16,AU23,AU26,Left,Lower Right,Upper

Fig. 3: Confusion Matrices for Entire Data Set



Correctly Perceived Data Set:

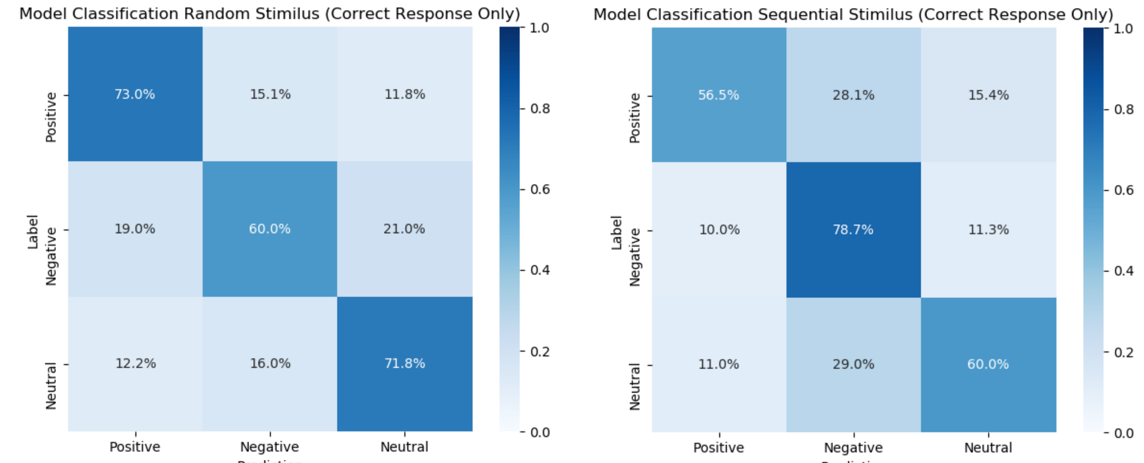
Table V: Classification Rate of Decision Tree on Correctly Perceived Data Set

Stimulus Order	Classification Rate of Decision Tree
Random (Correctly perceived data)	69.6%
Sequential (Correctly perceived data)	66.6%

Table VI: Selected AOI for Machine Learning on Correctly Perceived Dataset

Stimulus Order	Selected AOI
Random (All data)	AU1,AU2,AU6,AU9,AU12,AU15, AU20,AU23,AU26,Left,Lower,Right, Upper
Sequential (All data)	AU4,AU7,AU9,AU12,AU16,AU20, AU23,AU26,Left,Right,Upper

Fig. 4: Confusion Matrices for Correctly Perceived Data Set



DISCUSSIONS

Decision Tree vs. Participant

- With sequential video stimulus Participants have higher classification rate than decision tree
- With random video stimulus Participants have lower classification rate than decision tree

Random vs. Sequential

- With both entire data set and only correctly perceived data set, random stimulus had a higher classification rate than sequential stimulus

Entire data set vs. Perceived data set

- With both random and sequential stimulus, perceived data set had higher classification rates than entire dataset

CONCLUSION

Human emotions are subtle and can be expressed through a variety of different audio visual channels. In this study we focused on the human perception of facial expressions and movements as a channel to classify emotions. We studied eye fixation while participants watched a series of video stimulus of actors expressing emotions. Using a decision tree and statistical tests, the analysis showed that fixation on facial movements was an effective channel for automatic emotion recognition. We discovered that with random video stimulus, the eye fixation on AOI has better machine learning results and there are more AOI with significantly different means than sequential stimulus. Also when using only participants correctly perceived data, the machine learning models have better results and less confusion between the emotions.

FUTURE WORK

- Study difference in eye fixation and gaze data between audio and muted video stimulus
- Collect more data and explore deep learning options
- Develop new models based off of different eye gaze data besides fixation time in AOI such as gaze sequence.

REFERENCES

[1] A. A. T. Danisman, "Emotion classification of audio signals using ensemble of support vector machines," Lecture Notes in Computer Science, vol 5078. Springer, 2008
[2] H. L. N. Y. L. W. Qi Su, Fei Chen, "Multimodal emotion perception in children with autism spectrum disorder by eye tracking study," 8 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES),2018.
[3]W F P Ekman, "Facial action coding system: A technique for the measurement of facial movements" Consulting Psychologists Press, Palo.
[4] . C. Gaganjot Kaur, "Improved j48 classification algorithm for the prediction of diabetes," International Journal of Computer Applications (0975 – 8887) Volume 98 – No.22, July 2014 , 2014

ACKNOWLEDGEMENT

This research was supported by the National Science Foundation (NSF) Research Experiences for Undergraduates (REU) program. We would like to thank all of the mentors and research fellows at the New York Institute of Technology who have provided their helpful insight and expertise that greatly assisted with our research.
This project is funded by National Science Foundation Grant No. CNS-1852316 and New York Institute of Technology.