2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10340178

Emotion differentiation through features of eye-tracking and pupil diameter for monitoring well-being. *

Mackenzie L. Collins, *Member, IEEE,* T. Claire Davies, *Member, IEEE*

***Abstract*— Emotions are an important contributor to human self-expression and well-being. However, many populations express their emotions differently from what is considered "typical". Previous literature has indicated a possible relationship between emotion and eye-movement. The objective of this paper is to further explore this proposed relationship by identifying specific features of eye-movement that relate to six emotion categories: joy, surprise, indifference, disgust, sadness, and fear. Features of eye-movement are extracted from measurements of pupil diameter, saccades, and fixations. These measurements are collected as participants view images from the International Affective Picture System, a validated image deck used to evoke known levels of pleasure, arousal, and dominance. Example features of eye-movement measurements such as pupil diameter include maximum or minimum values, means, and standard deviations. Statistical analyses indicate that the extracted features of eye-tracking in this paper can identify fear and sadness with relative accuracy, while more work is needed to differentiate among joy, indifference, disgust, and surprise. Future work aims to understand differences between typically developing populations such as the individuals included in this analysis, and clinical populations such as individuals with cerebral palsy.**

***Clinical Relevance*— This pilot study suggests a link between emotion and features of eye-movement. The information will later be used to develop assistive communication devices that better meet the self-expression needs of individuals with motor and communication challenges.**

I. Introduction

Emotions play an integral role in our daily interactions and are important contributors to overall well-being. While emotions are a key aspect of human self-expression, not everyone conveys their emotions in the same manner. For instance, an individual with cerebral palsy may have complex communication needs resulting from difficulties producing clear speech due to motor control impairments. Previous literature has linked emotion to eye-movement signals, including changes in eye-movement features such as mean pupil diameter or maximum saccade velocity [1]. As acquisition of these eye-related signals does not require direct physical contact, it offers an opportunity to provide a method of expression to persons unable to communicate in a manner deemed "typical" by the general population (e.g., through speech or facial expressions) [1].

Emotion is modelled in one of two ways; dimensionally or discreetly [2]. Lang proposed the valance-arousal model in 1995, which consists of a two-dimensional space where various intensities of valence (i.e., unpleasant to pleasant) and arousal (i.e., passive to active) can describe the continuum of emotions in the human experience [2,3]. While this model can distinguish between positive and negative emotions successfully, it struggles with similar emotions such as anger or fear [3,4]. Mehrabian proposed a third dimension, dominance (i.e., submission to dominance), which aids in differentiating between such emotions [5]. This three-dimensional emotion model is commonly referred to as the pleasure-arousal-dominance or PAD model [5].

Discrete emotion models make use of terms such as anger, fear, and disgust to describe emotions in a colloquial manner that is more easily translatable to the general population [2]. A common discrete model of emotion was proposed by Ekman which describes human expression as having six basic emotions: joy, surprise, sadness, fear, disgust, and anger [6].

In current literature, physiological signal modalities used for emotion recognition include electroencephalography (EEG), galvanic skin response (GSR), and eye-tracking systems [1,7-9]. Unlike behavioral responses, physiological signals are under the purview of the autonomic nervous system (ANS) and therefore are not under conscious control [10]. This makes physiological signals preferable for emotion recognition systems as these signals contain less bias than those of behavioral responses [10]. However, these signals which provide insight into the innate response of the ANS present unique ethical challenges. The act of sharing emotions with others should remain voluntary. An individual's agency, or ability to act on one's will, is removed if emotions are automatically interpreted from signals that they are unable to control or mask. Therefore, privacy will be an important ethical consideration in the development of assistive technologies for emotion recognition.

Recent literature has explored the benefits of both unimodal and multimodal approaches for collecting and interpreting physiological signals. Lu et al. has examined the hybrid use of EEG and eye-tracking for emotion recognition and found that algorithm accuracy increased with the use of data from both modalities [11]. Guo et al., reported emotion recognition through machine learning with an accuracy of 59.66% when using eye-tracking alone, but when additional modalities (i.e., EEG and eye images) were added, accuracy increased to 79.63% [12].

While research has suggested that multimodal systems may be beneficial for accuracy, it is important to consider the needs of various stakeholders when designing a system. For instance, many individuals with cerebral palsy prefer devices that do not touch their body. Unlike EEG or GSR, various eye-

*Research supported by Vanier Canada Graduate Scholarships, NSERC
M. L. Collins and T. C. Davies are with the Mechanical and Materials Engineering Department at Queen's University, Kingston, ON, K7L 3N6, CA (corresponding author e-mail: 15mcl03@queensu.ca).

tracking technologies do not require direct physical interaction with the subject. This is a contributing factor to eye-tracking technology's continual study in the context of emotion recognition [11,12,14]. Our long-term aim is to design a device that can detect emotions while relying only on features obtained from a single, non-contact modality: eye-tracking.

Various features of eye-tracking have been used for emotion recognition in machine learning algorithms [11-14], including mean pupil diameter, standard deviation of pupil diameter, fixation duration (i.e., time spent focusing on a single point), mean saccade amplitude (i.e., movement between fixations), and maximum saccade velocity. The statistical analysis approach in this paper was inspired by Tarnowski et al. who used analysis of variance (ANOVA) to compare each feature of eye movement between three classes of emotions to inform feature selection for their machine learning models [14]. Tarnowski et al. found that variables such as average saccade amplitude, average fixation duration, and average pupil diameter demonstrated statistical differences among emotion classes [14]. One challenge when using machine learning algorithms for emotion recognition systems is the lack of diversity in training datasets. Many studies recruit young adult, typically developing (TD) individuals [13,14]. Use of these populations inhibits generalizability to individuals who differ from this population, despite their potential to benefit from the device.

Prior literature has primarily focused on positive, neutral, and negative emotions without differentiating within those categories [11,15]. Recently, researchers have begun expanding their emotion recognition work, categorizing more specific emotions, such as happiness, sadness, neutrality, fear, and anger [12-14]. The objective of this research is to evaluate accuracy of eye-tracking features at differentiating among six discrete emotion categories independent of other modalities. The results of this work will inform the design of machine learning algorithms used in assistive technologies for emotion recognition.

II. Methods

A. Stimuli

The International Affective Picture System (IAPS) was used to induce emotion in participants [16]. IAPS is a standardized collection of picture decks containing color images across a wide range of semantic categories [16]. These images are emotionally evocative, internationally accessible, and have been used across several studies to evoke emotion [16]. IAPS was selected for this study as it includes a validated photo deck for both adult and pediatric populations, and each image contains a mean and standard deviation score from one to nine for each of the three PAD dimensions of emotion: pleasure, arousal, and dominance [16].

B. Participants

Sixteen participants (n=16 (9F,7M), age=23.1(1.78)) were included in this study. Ethics was granted by the Queen's University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board (HSREB). Participants signed informed consent prior to participating in this study. Participants had self-reported normal or corrected-to-normal vision and they did not identify as members of a population with severe visual impairment, severe motor and communication impairment, or cerebral palsy. Participants were warned that some images may be of a sensitive nature and that they could ask to stop the procedure at any time should they feel overly uncomfortable.

C. Protocol

In a controlled laboratory environment, the Tobii TX300 Eye-Tracker mounted to the Tobii integrated 23" TFT monitor (1920 x 1080 pixel) was used to collect eye-movement signals. Participants were seated with their eyes approximately 65 cm from the center of the screen. Blinds were closed but ceiling lights remained illuminated to better simulate an everyday environment. Participants completed a nine-point calibration using the Tobii Pro Lab Software (Version 1.152). Participants were instructed to remain as still as possible while blinking normally for the duration of the protocol.

Images selected from the IAPS deck validated for individuals aged 7-9 years were used for all participants, as younger individuals met the inclusion criteria for this work. However, all individuals included in the current analysis are above the age of seventeen. After removal of images researchers felt were too sensitive for the purpose of the current study (removed images were agreed upon by a group of three researchers), fifty-two images remained. Each participant viewed a total of thirty images from this deck, randomly selected and randomly ordered.

Following calibration, the screen showed a message in black text on a grey background that asked the participant's verbal permission to begin. Preceding each image, a grey slide with the words "Get ready to rate the next image…" were shown on the screen for five seconds to normalize the pupil diameter between emotion-evoking images. One image from the deck on a grey background was then shown for six seconds, as recommended by the IAPS manual [16]. This was repeated for thirty images total.

D. Creating Emotion Categories

Zhang et al. introduced a method for interpreting pleasure, arousal, and dominance scores as a discrete model of emotion [17]. Using Bradley et al.'s mean and standard deviation measurements for PAD scores relating to each affective term [18], an ellipsoid is created in three-dimensional space. For a given emotion, the mean PAD scores represent the center of the ellipse, while the standard deviations in each direction define the semi-axes [17,18]. To categorize the stimuli from IAPS into emotion categories, the PAD scores provided by the technical manual were plotted and compared to the ellipsoids to determine which centroid (mean emotion) was the shortest Euclidean distance from the image's PAD score. The closest centroid of an ellipsoid determined the image's emotion category. If the image's PAD score did not fall within the bounds (i.e., standard deviations) of the ellipsoid, in the current analysis, it was still categorized by the closest centroid.

E. Data Analysis

Tobii Pro Lab (Version 1.152) was used for data collection, initial filtering (Tobii Pro I-VT Fixation Filter), and data analysis. Additional feature extraction was completed in MATLAB (Windows, Version R2022a). Statistical analysis was complete in IBM SPSS Statistics (Windows, Version

22.0). Assumption testing was conducted on each feature to ensure both normality (kurtosis and skew approximately between -1.00 to +1.00) and homogeneity of variance (Levine's test). All variables passed Levine's test; however, several variables demonstrated positive skew and underwent log-transformation to correct for normality. A one-way ANOVA was used to compare the six emotion classes for each feature of eye-movement [14]. A Bonferroni correction was used to determine the significance level for this study given 18 ANOVAs were completed. Resultantly, the new significance value for this study is $\alpha = 0.05/18 = 0.003$. A feature is a single value such as a maximum, minimum, or mean that is extracted from a physiological measurement, such as pupil diameter, saccades, and fixations. A Tukey post-hoc test was used to further explore the differences among emotion classes.

III. RESULTS

A. *Categorizing stimuli into affective groups*

Bradley et al.'s reported mean and standard deviation PAD ratings for various affective terms were used to create the ellipsoids seen in Figure 1 as described by Zhang's approach [17,18]. Each individually plotted data point (asterisks) represents the mean PAD score for an IAPS image [16]. The closest center of an ellipsoid was used to categorize each image into one of the seven categories: indifference, joy, surprise, sadness, fear, disgust, or anger. The results of this procedure are illustrated in Fig. 1. In this study, none of the images selected were categorized as evoking "anger", and therefore this emotion was not assessed in the subsequent analysis, leaving six emotion categories total.

B. *ANOVA*

To compare the features' ability to differentiate amongst the six emotion categories, a one-way ANOVA with a Bonferroni correction was completed for each feature of eye movement. Each test compared a feature of eye-movement or pupil diameter between the six classes of emotion. Three of these metrics produced statistically significant results. Table 1 indicates the results from this ANOVA. Of the statistically significant results, one related to features of saccades, and two related to features of pupil diameter.

C. *Tukey Post Hoc Test*

A Tukey post hoc test was conducted on the statistically significant features identified by the ANOVA test. Results in Table 2 indicate that the selected features can differentiate sadness and fear from other emotions, while differentiating amongst disgust, indifference, joy, and surprise remains a challenge.

IV. DISCUSSION

Statistically significant results were obtained for various features of eye-movement, compared across the six emotion classes. These differences indicate that emotion may be related to eye-movement. Features relating to pupil diameter and saccades provided statistically significant results, while features related to fixations did not. These results align with those of Peng et. al, who indicated pupil size as a "powerful physiological indicator" in emotion recognition scenarios [19]. Current literature has recognized the role of pupil diameter in emotion recognition and particularly its use in differentiating

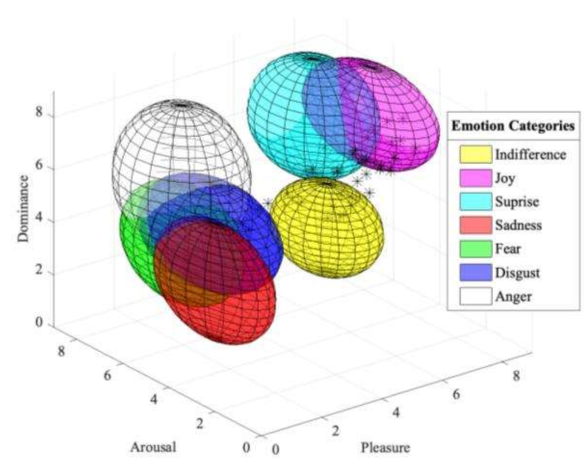

Figure 1. Categorizing IAPS images (stimuli) into affective groups (ellipsoids).

between positive and negative emotions [11-13,19]. Aracena et al. also discussed the benefits of using eye-tracking metrics such as pupil diameter for emotion recognition due to its cost effectiveness and practicality over other technologies [13]. The standard deviation of saccade velocity performed well in the identification of fear and sadness. Notably, the standard deviation of peak saccade velocity demonstrated a significance of <0.001, suggesting it has great potential for emotion differentiation.

Post hoc analysis showed that while current metrics can identify sadness and fear, more research is required as to the other four categories (disgust, indifference, joy, and surprise) for reliable emotion differentiation.

V. LIMITATIONS

There are several limitations in the current study. First, the small sample size may be impacting the significance of

TABLE 1: ONE-WAY ANOVA RESULTS

Eye Feature	**Emotion Category**	
	F (5,465) =	***p***
Mean fixation duration	0.396	0.85
Number of whole fixations	1.108	0.36
Number of saccades	0.802	0.55
Minimum duration of whole fixations	0.537	0.75
Time to first whole fixation	0.361	0.88
Duration of first whole fixation	2.964	0.01
Mean peak saccade velocity	2.539	0.03
Minimum peak saccade velocity	1.617	0.15
Max peak saccade velocity	1.709	0.13
STD of peak saccade velocity*	4.644*	<0.001*
Mean saccade amplitude	2.311	0.04
Maximum amplitude of saccades	2.199	0.05
Mean right pupil diameter	2.459	0.03
Mean left pupil diameter	2.657	0.02
STD of left pupil diameter*	4.983*	<0.001*
STD of right pupil diameter*	4.152 *	0.001*
Max left pupil diameter	2.953	0.01
Max right pupil diameter	2.427	0.04

* Indicates statistical significance

TABLE 2: TUKEY POST-HOC ANALYSIS RESULTS

Eye Feature	p-Values between emotion classes								
	Disgust Sadness	***Fear Sadness***	***Indifference Sadness***	***Joy Sadness***	***Surprise Sadness***	***Disgust Fear***	***Indifference Fear***	***Joy Fear***	***Surprise Fear***
STD of peak saccade velocity	0.003*	<0.001*	<0.001*	<0.001*	0.023	-	-	-	-
STD of left pupil diameter	-	-	-	-	-	0.017	0.002*	<0.001*	0.022
STD of right pupil diameter	-	-	-	-	-	0.040	0.002*	0.001*	-

* Indicates statistical significance

results. Second, the randomization of image selection led to each image having a different number of respondents. Similarly, not every emotion category contained the same number of images. This resulted in additional variation among the group sizes.

As a conservative approach, a one-way ANOVA was completed between emotion categories, with a Bonferroni adjustment made to the significance level to account for the completion of multiple ANOVAs. The Bonferroni correction aims to reduce the chances of obtaining false-positive results (Type I errors), as the probability of a significant result occurring due to chance increases as the number of completed tests increases. This in turn elevates the chance of false-negative results (Type II errors), meaning this conservative approach may be missing valuable information, and other features may play a role in emotion differentiation despite not being recognized as significant in the current study. The use of this statistical test was both a strength and a limitation of this work, in that it provided an estimate of the most significant individual features, while likely overlooking the value in others. The method of applying one-way ANOVAs also does not account for the potential error introduced by different respondents and uneven number of images per emotion category.

For future work, focus should be placed upon pupil diameter and saccade features for differentiating among emotions. Advanced statistical, along with machine learning techniques will be considered to further emotion recognition accuracy using features of eye-movement alone.

VI. CONCLUSION

Communicating emotion is a challenge for many populations with motor or communication impairments. The eye-tracking features analyzed herein will help inform the development of emotion-recognition and communication systems. Eye movement features including those associated with pupil diameter and eye saccades were able to identify fear and sadness. Future work should focus on making use of advanced statistics and machine learning techniques to better differentiate among emotions. Clinical populations for whom this system is necessary will be involved in future work as they have the potential to benefit greatly from affective-computing technologies. Variations among clinical populations and TD persons will need to be compared prior to the development of an emotion-recognition system.

VII. REFERENCES

[1] J. Z. Lim, J. Mountstephens, and J. Teo, "Emotion Recognition Using Eye-Tracking: Taxonomy, Review and Current Challenges", Sensors, vol. 20, no. 8, p. 2384, 2020.

[2] W. K. Chen, Linear Networks and Systems (Book style).Belmont, CA: Wadsworth, 1993, pp. 123–135.

[3] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, and X. Yang, "A Review of Emotion Recognition Using Physiological Signals", Sensors, vol. 18, no. 7, p. 2074, 2018.

[4] P. J. Lang, "The emotion probe: Studies of motivation and attention.," American Psychologist, vol. 50, no. 5, pp. 372–385, 1995.

[5] A. Mehrabian, "Comparison of the PAD and PANAS as models for describing emotions and for differentiating anxiety from depression," Journal of Psychopathology and Behavioral Assessment, vol. 19, no. 4, pp. 331–357, 1997.

[6] P. Ekman, "An argument for basic emotions," Cognition and Emotion, vol. 6, no. 3-4, pp. 169–200, 1992.

[7] Z. Ahmad and N. Khan, "A survey on physiological signal-based emotion recognition," Bioengineering, vol. 9, no. 11, p. 688, 2022.

[8] N. S. Suhaimi, J. Mountstephens, and J. Teo, "EEG-based emotion recognition: A state-of-the-art review of current trends and opportunities," Computational Intelligence and Neuroscience, vol. 2020, pp. 1–19, 2020.

[9] A. Goshvarpour, A. Abbasi, and A. Goshvarpour, "An accurate emotion recognition system using ECG and GSR Signals and matching Pursuit Method," Biomedical Journal, vol. 40, no. 6, pp. 355–368, 2017.

[10] J. Tao and T. Tan, "Affective Computing: A Review," in Affective Computing and Intelligent Interaction: Springer Berlin Heidelberg, 2005, pp. 981-995.

[11] Y. Lu, W. L. Zheng, B. Li, and B. L. Lu, "Combining eye movements and EEG to enhance emotion recognition,", 2015, pp. 1170-1176.

[12] J. J. Guo, R. Zhou, L. M. Zhao, and B. L. Lu, "Multimodal Emotion Recognition from Eye Image, Eye Movement and EEG Using Deep Neural Networks," in 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2019, July 23, 2019 - July 27, 2019, Berlin, Germany, 2019: Institute of Electrical and Electronics Engineers Inc., in Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, pp. 3071-3074, doi: 10.1109/EMBC.2019.8856563.

[13] C. Aracena, S. Basterrech, V. Snael, and J. Velasquez, "Neural networks for emotion recognition based on eye tracking data," 2015 IEEE International Conference on Systems, Man, and Cybernetics, 2015.

[14] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, "Eye-Tracking Analysis for Emotion Recognition," Computational Intelligence and Neuroscience, vol. 2020, pp. 1-13, 2020, doi: 10.1155/2020/2909267.

[15] M. Soleymani, M. Pantic, and T. Pun, "Multimodal Emotion Recognition in response to videos," IEEE Transactions on Affective Computing, vol. 3, no. 2, pp. 211–223, 2012.

[16] M. B. Peter Lang, Bruce Cuthbert, "International affective picture system (IAPS): Affective ratings of pictures and instruction manual.," University of Florida, Gainesville, FL, Technical Report A-8., 2008.

[17] S. Zhang, Z. Wu, H. M. Meng, and L. Cai, "Facial expression synthesis based on emotion dimensions for affective talking avatar," Modeling Machine Emotions for Realizing Intelligence, pp. 109–132, 2010.

[18] M. Bradley, P. J. Lang, "Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings.," University of Florida, Gainesville, FL, USA, Technical Report C-1., 1999.

[19] R. Peng, A. Barreto, G. Ying, and M. Adjouadi, "Affective Assessment by Digital Processing of the Pupil Diameter," IEEE transactions on affective computing, vol. 4, no. 1, pp. 2-14, 2013, doi: 10.1109/T-AFFC.2012.25.