# DiVA: Exploring the Usage of Pupil Diameter to Elicit Valence and Arousal

**Khaled Kassem, Jailan Salah, Yasmeen Abdrabou, Mahesty Morsy, Reem El-Gendy, Yomna Abdelrahman*, Slim Abdennadher**
German University in Cairo, Egypt, {firstname.lastname}@guc.edu.eg
*University of Stuttgart, HCI Group, Germany, {firstname.lastname}@vis.uni-stuttgart.de

## ABSTRACT
Most of the typical digital systems are not fully aware of the users' affect states. Adapting to the users' state showed great potential for enhancing user experiences. However, most approaches for sensing affective states,specifically arousal and valence, involve expensive and obtrusive technologies, such as physiological sensors attached to users' bodies. This paper present an indicator of the users' affect based on eye tracking. We use a commercial eye tracker to monitor the user's pupil size to estimate their arousal and valence in response to videos of different content. To assess the effect of different content (namely pleasant and unpleasant) influencing the arousal and valence on the pupil diameter, we conducted a user study with 25 participants. The study showed that different content of videos affect the pupil diameter, thereby giving an indicator about the user's state. We provide empirical evidence showing how to unobtrusively detect changes in users' state. Our initial investigation gives rise to eye-based user's tracking, which introduces the potential of new applications in the field of affect-aware computing.

### ACM Classification Keywords
H.5.2. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

### Author Keywords
Arousal; Valence; Eye Tracker; Pupil Diameter.

## INTRODUCTION
Affect-aware systems aim to sense users' internal states and to adapt their interface and behavior accordingly. Such systems offer opportunities to tailor activities in areas such as work, health care, education, and even gaming. For example, affect-aware systems can enhance the quality of education by responding appropriately to a student's emotional feedback. Accurate reliable automatic detection of emotions is a necessary first step in any affect-aware system.



*MUM 2017,* November 26–29, 2017, Stuttgart, Germany

 ISBN 978-1-4503-5378-6/17/11…$15.00
DOI: https://doi.org/10.1145/3152832.3152836

Despite over 50 years of work in the area, how to sense affect state in a robust, accurate, timely, and unobtrusive way is still an open challenge. Affect state has been measured traditionally by subjective self-reporting. For instance, using the Self-Assessment Manikin (SAM)[6] is a common example of the first category, where participants are asked to report their own state. Other approaches include observing the human. Human emotions tend to show themselves in many ways, the first of which is human behavior. Humans engaged in a conversation often exhibit nonverbal cues such as body language, facial expression, temperature [2] and tone of voice [9]. Extensive research has been done to deduce the emotional state of individuals from facial expressions, tone of voice [9], or text-based [21] interaction.

Recently, a lot of work has aimed at studying physiological indicators of human emotions [1, 2, 10], such as heart rate and skin temperature [17]. Intuitively, physiological changes can indicate the onset of certain emotions, e.g. a dangerous situation that triggers a sense of fear causes a fight-or-flight response in which heart rate is elevated. Using the appropriate bio-sensors to measure and record such involuntary physiological changes that occur in response to emotions is, therefore, an appropriate way of automatically detecting and identifying emotional states. This area of inquiry has become more popular in recent years, since technological advances that made such sensors smaller, less obtrusive, and more efficient.

Eye tracking is a strong candidate for state measuring. Eye trackers are unobtrusive and operate in real time. Previous research has shown that eye features reveal different aspects of our internal states, including attention [15], affect [19] and stress [12]. Further, advances in miniaturization and mass production have continuously brought down the prices of these devices. Consumer-grade trackers are readily available, and make monitoring affect states on a large scale feasible.

In this paper, we present a method for estimating users' arousal and valence based on pupil size using a commercial wearable eye tracker. Our method works with off-the-shelf hardware and is applicable for ubiquitous computing environments. It estimates the arousal and valence state of the user and opens up new opportunities for deployments of affect-aware technologies.

In this work, we advance the state-of-the-art of automatic affect state estimation through the following contributions:

- We propose a method for estimating the arousal and valence using a single sensor as well as a single metric.
- We demonstrate the validity of our metric through a user study, showing that the estimate for the user state strongly correlates with the content difference in a naturalistic task (watching videos of different genres).

## RELATED WORK

Awareness of user's emotional state is an ongoing challenge for HCI research. Our work builds on strand of prior research on utilizing wearable eye tracking for affect state estimation.

### Affect State Estimation

Observing users with the aid of sensors can give cues about their affect states. Affect recognition has been involved in vast number of emotional aware applications and domains e.g. tutoring systems, gaming, affective computing, health care, customized websites, call centers and adaptive systems[14].

Different approaches have been introduced to estimate affect states ranging from using facial expressions [8], facial temperature [17], eye movements [18], Galvanic Skin Response (GSR) [10], brain signals [20], and Electrodermal activity (EDA) [16].

One of the main advantages of physiological sensors, such as GSR, Blood Volume Pressure (BVP), and pupil diameter is that they are controlled by the Autonomic Nervous Systems (ANS). The nerves that control the size of the pupil react to emotions changing the pupil size involuntarily, which ensures accurate detection of emotions[19]. On the other hand, a major limitation of current physiological approaches is the price of the sensors as well as the need for sensors to be in direct contact with the user.

Eye trackers showed the potential to be one of the reliable biosensors in detecting emotions. Eye trackers are now commercially available in terms size and cost. Moreover, they could be used in a wearable and remote forms. Several studies explored the usage of eye trackers to estimate emotional content based on the pupillary size difference.

#### *Visual Stimuli*

Previous research used images to elicit emotions [3, 7, 15, 13]. They showed that pupillary diameter becomes larger on emotional arousing stimuli regardless of whether they were pleasant or unpleasant. Moreover, [7] showed that pupil's response during picture viewing reflects emotional arousal associated with increased sympathetic activity.

#### *Audio Stimuli*

Other studies used audio stimuli [4, 5, 19]. They investigated how they relate to pupil size variation with different stimuli. Results showed that pupil size was significantly larger after both negative and positive than neutral stimulation.

In summary, most of the research used different stimuli to explore the effects of emotional arousal and valence on different eye activity features. The results supported the hypothesis that pupillary size is an indicator of individuals' emotional states as it dilates with emotional states regardless of whether it is pleasant or unpleasant emotions.

## EYE TRACKING FOR AROUSAL AND VALENCE ESTIMATION

Limited research has been conducted to infer valence from pupil size. Tavakoli et al. [11] were successful to elicit valence, they combined 10 eye movement features to recognize valence caused by visual stimuli.

In this work, we aim to investigate the emotional valence using one metric only, namely, the pupil diameter size with a commercial sensor using multi-sensory stimuli (video clips). This stimuli was chosen because it is closer to real life situations that elicit different emotions. The commercial eye tracker was chosen because one of the objectives of this study is to assess the feasibility of the usage of commercial sensors to be embedded in systems that are applied not only in experimental environments. In summary, in this work, we focus on one main research question:

- Can we distinguish arousal and valence using a commercial wearable eye tracker? More specifically, do the changes in pupil diameter correlate with valence as it correlates with arousal? (**RQ**)

## STUDY: CORRELATING VALENCE AND AROUSAL WITH PUPIL DILATION

To test our hypothesis of the ability of eye trackers to derive arousal and valence states based on pupil dilation percentage, we conducted a user study in which we recorded the pupil diameter of participants' in two activity states: 1) Relaxing as the baseline and 2) watching videos with different content types.

### Design

We applied a repeated-measures design, all participants were exposed to all conditions. We studied the effect of the tasks on the pupil diameter. For the baseline, we asked the participants to relax. For the video watching task we provided 5 videos with 2 different content types, namely pleasant and unpleasant content:

1. Pleasant:
   (a) "Smile" by Sia [1],
   (b) "Hoverboard in Real Life! In 4K"[2],
   (c) "The Nights" by Avicii[3],
2. Unpleasant:
   (a) "The Last of Us" [4],
   (b) and a death scene from Disney's "The Lion King" [5].

[1] https://www.youtube.com/watch?v=IKjJ6DQF7xY
[2] https://www.youtube.com/watch?v=gMaDhkNJA2g
[3] https://www.youtube.com/watch?v=UtF6Jej8yb4
[4] https://www.youtube.com/watch?v=DidjobKweCs
[5] https://www.youtube.com/watch?v=4Jo9OQfXstM

We chose these content types because of their presumed differences in affect. Moreover, we aimed to evaluate the estimation of the arousal and valence using naturalistic content, rather than controlled lab stimuli.

To confirm our presumed content types, 28 volunteers were asked to rate the emotional impact of the videos using the scale [6]. The videos were evaluated on two aspects, both on a numerical scale from 1 to 9: cheerfulness (1 being saddest/least cheerful, 9 being happiest/most cheerful), and emotional arousal (1 being least arousing, 9 being most arousing). Based on the collected rating our presumed classification was confirmed.

Additionally, We collected subjective rating from the participants during the study to reconfirm our categorization of the videos used. To overcome the order effect, the order of the tasks was counter-balanced using a Latin Square.

**Figure 1. Setup: eyetracker, display and user performing the task.**

### Apparatus
Our experimental setup consisted of the Pupil wearable eye tracker [6]. The gaze accuracy and precision are 0.60 and 0.08 degrees respectively with frame rate of 120 Hz. A Lenovo Y510P with a screen resolution of 1366x768 was used to display the videos for the participants. The eye tracker uses USB as power source as well as to transfer data. The participants were asked to look at the screen placed at 0.8m from the participant as shown in figure 1.

### Data Recordings
We utilized the *Pupil Capture* and *Pupil Player* [7]. *Pupil Capture* was used for recording videos from the eye camera. Pupil Player was used to extract data from the recorded videos recorded by *Pupil Capture* to CSV files for later analysis.

### Participants and Procedure
We recruited 25 participants (7 females) with an average age of 21.74 years ($SD = 2.35$) using university mailing lists. None of the participants were wearing eye glass. Participants were students of different majors. None of the participants had any previous experience with eye trackers.

After arriving in the lab, participants filled in their demographics and signed a consent form and received an explanation of the purpose of the study. First, participants were asked to relax for 4 minutes.Next, we asked participants to watch the set of videos, each for 2 minutes followed by Grey screen for 5 seconds. The order of the tasks was counter-balanced using Latin-square. After each of the video watching task, they were asked to rate the video by filling in SAM scale [6].

The study took approximately 30 minutes. During the entire experiment, we recorded the mean dilation value and the confidence level from the pupil responses. The experiment was conducted with a maintained screen brightness to isolate any influence on the recorded pupil diameter.

## RESULTS
We analyzed the effect of the video content and how they influence the arousal and valence on the recorded pupil diameter. We used a single metric as our dependent variable: variation in pupil diameter. We defined the pupil diameter change as the difference between the mean diameter during the baseline recording and the mean diameter during the performance of the task.

### Video Rating
Regarding the SAM ratings, a paired-samples t-test was conducted to compare the mean value between the valence for the pleasant and unpleasant stimuli.

There was a significant difference in the mean values between the pleasant and unpleasant stimuli (M=3.94, SD= 1.59) with $p < 0.001$ for all videos. These results suggest that the SAM results are reliable and can be taken into consideration while choosing the stimulus.

### Effect of Arousal on the Pupil Diameter
Regarding the arousal testing results, a paired-samples t-test was conducted to compare pupil dilation in neutral and emotional conditions. It compares between pupil dilation in the baseline phase and the phases of watching the pleasant/unpleasant videos. There was a significant difference in the scores for the dilation percentage of the baseline and each emotional video of the five ones chosen (Table 5.2) with $p < 0.001$ for all pairs.

These results suggest that watching the emotional videos does have an effect on pupil dilation regardless if they are pleasant or unpleasant. Specifically, our results suggest that when users get affected by arousal stimuli, they have their pupil size dilated.

| Video | Arousal (Mean::SD) | | Valence (Mean::SD) | | Pupil Diameter (Mean::SD) | |
|---|---|---|---|---|---|---|
| Smile | 6.04 | 2.19 | 6.63 | 1.01 | 14.7 | 11.1 |
| The nights | 7.11 | 1.93 | 7.25 | 1.48 | 13.8 | 9.7 |
| Hoverboard | 5.92 | 1.64 | 5.88 | 1.11 | 13.5 | 12.2 |
| The last of us | 7.32 | 1.96 | 2.8 | 1.15 | 22.9 | 13.6 |
| Lion King | 6.53 | 2.66 | 3.2 | 3.91 | 22.9 | 13.2 |

**Table 1. Mean and Standard Deviation of the Subjective Ratings of the Videos and the Pupil Size Change Relative to the Baseline.**

[6] https://pupil-labs.com/pupil/

[7] https://github.com/pupil-labs/pupil/releases/tag/v0.9.13/

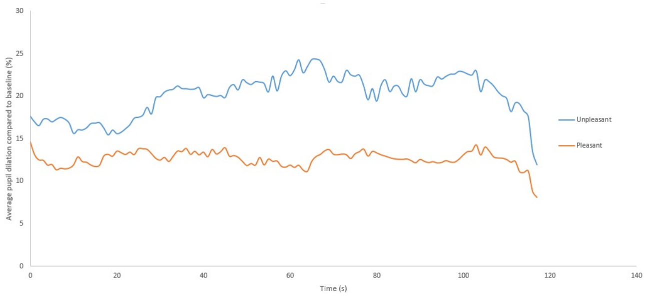


**Figure 2. Average percentage of pupil dilation when exposed to pleasant (red) and unpleasant stimuli (blue).**

### Effect of Valence on the Pupil Diameter

A one-way repeated measures ANOVA was conducted to compare the effect of the videos on pupil dilation percentage in pleasant and unpleasant conditions. There was a significant effect of video type Wilks’ Lambda = 0.29, F (4,25) = 14.87, p =0.00. Paired samples t-tests were used to make post hoc comparisons between different video types. The first two paired samples t-test indicated that there was a significant difference in the pupil dilation percentage between the first pleasant video "Smile" (M= 14.7, SD=11.1), and both unpleasant videos (M1=22.9, SD1=13.6) t(24)= -5.33, (M2=22.9, SD2=13.2) t(24)= -5.18, with p = .000 for both tests.

A second pair of paired samples t-test indicated that there was a significant difference in the pupil dilation percentage between the second pleasant video "The Nights" (M= 13.8, SD=9.7), and both unpleasant videos t(24)= -5.8, and t(24)= -5.6 with p = .000 for both tests.

A third pair of paired samples t-test indicated that there was a significant difference in the pupil dilation percentage between the third pleasant video "Hoverboard" (M= 13.5, SD=12.2), and both unpleasant videos t(24)= -7.2, and t(24)= -6.8, with p = .000 for both tests. The paired sample t-tests between each two videos from the same category had no statistical significance difference in pupil dilation (p>0.05), therefore our **RQ** is positively answered, where there is difference in pupil size upon valence.

### State Classification

In order to identify the different states, we needed to compute a threshold for the dilation values. We systematically tested the threshold values from the graph as shown in Figure 2. Finally, we calculated a general percentage to elicit emotions change. As shown in Figure 2 emotions can be triggered after 10% dilation from the baseline. In addition, if the eyes dilated more than 16%, then this can identify unpleasant emotions. Finally, pleasant emotions can be identified if the dilation falls between 10% and 15%.

## DISCUSSION

Informed by previous work, we hypothesized that change in arousal and valence would lead to a change in the participants’ pupil diameter. We tested the effects of different tasks and their content on pupil diameter changes. We elicited change in the affected state through a naturalistic task. We found that a change in video content leads to a change in the pupil diameter. We found an average increase in the diameter ranging between 10%-15% for pleasant content and above 16% for unpleasant content. The pupil diameter increases are correlated with the activation of the ANS, where the nerves that control the size of the pupil react to emotions changing the pupil size involuntarily. Our results answer **RQ** by showing that the pupil diameter varies differently in response to the valence state of the user.

## LIMITATION AND FUTURE WORK

In this work, we demonstrated how to use eye trackers to assess arousal and valence. Our findings suggest that commercial eye trackers are a suitable method for assessing individual’s affect state. However, our approach has its limitations. We acknowledge that we considered a controlled setup, where the brightness of the surroundings was kept constant. Also, we used a wearable eye tracker, exploring the usage of remote eye tracker would lead to touch-less, and unobtrusive method to estimate users’ state. Another limitation is the added effect of scene changes within individual clips on pupil size. However, more sophisticated approaches (e.g. using machine learning) could be used to consider the influence of the surrounding brightness change.

Further studies can be made using a wider pool of randomly sampled clips to obtain more generalizable results. Further investigation can be made into the effect of individual emotional triggers on the pupillary response. We aim to further validate the computed thresholds for pleasant and unpleasant states. In future research, we aim to evaluate the performance of using single metric as opposed to other multiple features for instance the 10 eye features as proposed by Tavakoli et al. [11].

## CONCLUSION

In this work we described our approach to derive users’ state based on pupil diameter using commercial wearable eye tracker. We investigated the effects of two different video content on pupil diameter changes. We observed significant changes in the pupil diameter upon the activation of the ANS due to a stimulus. Based on these observations, we proposed a technique for estimating and quantifying arousal and valence states using a single metric.

We envision that our results can be ported to be used by remote eye trackers, allowing a touch-less, and unobtrusive method to estimate users’ state. This would introduce wider scenarios of affect-aware applications including: adaptive systems, health care, education, and gaming providing novel and enhanced user experience.

## ACKNOWLEDGMENT

This work was supported by the German Federal Ministry of Education and Research, FeuerWeRR Grant No. 13N13481. This work was partly conducted within the Amplify project which received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement no. 683008) and received funding from the German Research Foundation within the SimTech Cluster of Excellence (EXC 310/2).

## REFERENCES


1. Yomna Abdelrahman, Mariam Hassib, Maria Guinea Marquez, Markus Funk, and Albrecht Schmidt. 2015. Implicit engagement detection for interactive museums using brain-computer interfaces. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. ACM, 838–845. DOI: http://dx.doi.org/10.1145/2786567.2793709
2. Yomna Abdelrahman, Eduardo Velloso, Tilman Dingler, Albrecht Schmidt, and Frank Vetere. 2017. Cognitive Heat: Exploring the Usage of Thermal Imaging to Unobtrusively Estimate Cognitive Load. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 33 (2017), 20. DOI: http://dx.doi.org/10.1145/3130898
3. Claudio Aracena, Sebastián Basterrech, Václav Snáel, and Juan Velásquez. 2015. Neural Networks for Emotion Recognition Based on Eye Tracking Data. In *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*. IEEE, 2632–2637. DOI: http://dx.doi.org/10.1109/SMC.2015.460
4. Areej Babiker, Ibrahima Faye, and Aamir Malik. 2013. Pupillary behavior in positive and negative emotions. In *Signal and Image Processing Applications (ICSIPA), 2013 IEEE International Conference on*. IEEE, 379–383. DOI:http://dx.doi.org/10.1109/ICSIPA.2013.6708037
5. Areej Babiker, Ibrahima Faye, and Aamir Malik. 2014. Differentiation of pupillary signals using statistical and functional analysis. In *Intelligent and Advanced Systems (ICIAS), 2014 5th International Conference on*. IEEE, 1–6. DOI:http://dx.doi.org/10.1109/ICIAS.2014.6869498
6. Margaret M Bradley and Peter J Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* 25, 1 (1994), 49–59. DOI: http://dx.doi.org/10.1016/0005-7916(94)90063-9
7. Margaret M Bradley, Laura Miccoli, Miguel A Escrig, and Peter J Lang. 2008. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* 45, 4 (2008), 602–607. DOI: http://dx.doi.org/10.1111/j.1469-8986.2008.00654.x
8. Carlos Busso, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, and Shrikanth Narayanan. 2004. Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proceedings of the 6th international conference on Multimodal interfaces*. ACM, 205–211. DOI: http://dx.doi.org/10.1145/1027933.1027968
9. Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz, and John G Taylor. 2001. Emotion recognition in human-computer interaction. *IEEE Signal processing magazine* 18, 1 (2001), 32–80. DOI: http://dx.doi.org/10.1109/79.911197
10. Mai ElKomy, Yomna Abdelrahman, Markus Funk, Tilman Dingler, Albrecht Schmidt, and Slim Abdennadher. 2017. ABBAS: An Adaptive Bio-sensors Based Assistive System. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2543–2550. DOI: http://dx.doi.org/10.1145/3027063.3053179
11. R Hamed, Adham Atyabi, Antti Rantanen, Seppo J Laukka, Samia Nefti-Meziani, Janne Heikkilä, and others. 2015. Predicting the valence of a scene from observers’ eye movements. *PloS one* 10, 9 (2015), e0138198. DOI: http://dx.doi.org/10.1371/journal.pone.0138198
12. Jennifer Healey and Rosalind Picard. 2000. SmartCar: detecting driver stress. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, Vol. 4. IEEE, 218–221. DOI: http://dx.doi.org/10.1109/ICPR.2000.902898
13. Cory S Inman. 2010. *Experience-related eye movements and pupillary responses reflect declarative memory for emotional and neutral pictures*. Ph.D. Dissertation. Emory University.
14. Alejandro Jaimes and Nicu Sebe. 2007. Multimodal human–computer interaction: A survey. *Computer vision and image understanding* 108, 1 (2007), 116–134. DOI: http://dx.doi.org/10.1016/j.cviu.2006.10.019
15. Shuhei Kawai, Hironobu Takano, and Kiyomi Nakamura. 2013. Pupil diameter variation in positive and negative emotions with visual stimulus. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*. IEEE, 4179–4183. DOI: http://dx.doi.org/10.1109/SMC.2013.712
16. Kyung Hwan Kim, Seok Won Bang, and Sang Ryong Kim. 2004. Emotion recognition system using short-term monitoring of physiological signals. *Medical and biological engineering and computing* 42, 3 (2004), 419–427. DOI:http://dx.doi.org/10.1007/BF02344719
17. Zhilei Liu and Shangfei Wang. 2011. Emotion recognition using hidden Markov models from facial temperature sequence. *Affective computing and intelligent interaction* (2011), 240–247. DOI: http://dx.doi.org/10.1007/978-3-642-24571-8_26
18. Yifei Lu, Wei-Long Zheng, Binbin Li, and Bao-Liang Lu. 2015. Combining Eye Movements and EEG to Enhance Emotion Recognition.. In *IJCAI*. 1170–1176.
19. Timo Partala and Veikko Surakka. 2003. Pupil size variation as an indication of affective processing. *International journal of human-computer studies* 59, 1 (2003), 185–198. DOI: http://dx.doi.org/10.1016/S1071-5819(03)00017-X
20. Alireza Sahami Shirazi, Markus Funk, Florian Pfleiderer, Hendrik Glück, and Albrecht Schmidt. 2012. MediaBrain: Annotating Videos based on Brain-Computer Interaction.. In *Mensch & Computer*. 263–272. DOI: http://dx.doi.org/10.1524/9783486718782.263

21. Shiv Naresh Shivhare and Saritha Khethawat. 2012. Emotion detection from text. *arXiv preprint arXiv:1205.4944* (2012).