Exploring Eye Activity as an Indication of Emotional States Using an Eye-Tracking Sensor

Sharifa Alghowinem[1,4], Majdah AlShehri[2],
Roland Goecke[3,1], and Michael Wagner[3,1]

[1] Australian National University, Canberra, Australia
[2] King Saud University, Riyadh, Saudi Arabia
[3] University of Canberra, Canberra, Australia
[4] Ministry of Higher Education, Kingdom of Saudi Arabia

Abstract. The automatic detection of human emotional states has been of great interest lately for its applications not only in the Human-Computer Interaction field, but also for its applications in psychological studies. Using an emotion elicitation paradigm, we investigate whether eye activity holds discriminative power for detecting affective states. Our emotion elicitation paradigm includes induced emotions by watching emotional movie clips and spontaneous emotions elicited by interviewing participants about emotional events in their life. To reduce gender variability, the selected participants were 60 female native Arabic speakers (30 young adults, and 30 mature adults). In general, the automatic classification results using eye activity were reasonable, giving 66% correct recognition rate on average. Statistical measures show statistically significant differences in eye activity patterns between positive and negative emotions. We conclude that eye activity, including eye movement, pupil dilation and pupil invisibility could be used as a complementary cues for the automatic recognition of human emotional states.

Keywords: Affective computing, eye tracking, emotion recognition.

1 Introduction

Affective computing – the study of automatic recognition of human emotional states and their utilisation in a computer system – has had much interest lately due to its multidisciplinary applications. For example, Human-Computer Interaction (HCI) is concerned with enhancing the interactions between users and computers by improving the computer's understanding of the user's needs, which include understanding the user's emotional state [23]. In the education field, understanding the affective state of a student could lead to more effective presenting style and improved learning [7]. A current interest is in the personalisation of commercial products, which could be enhanced by understanding the client's preference based on their mood [31]. Moreover, such understanding of the user's emotions could enhance other applications such as virtual reality and smart surveillance [29]. Such automatic recognition of emotions could also be useful to support psychological studies. For example, such studies could give a baseline for the emotional reaction of healthy subjects, which could be compared and used to diagnose mental disorders such as autism [14] or depression [1].

Eye-tracking applications cover several domains, such as psychology, engineering, advertising, and computer science [9]. As an example, eye-tracking techniques have been used to detect driver fatigue [15]. Moreover, cognitive load of a learner could be determined using eye-tracking and pupil measuring [4,18]. Moreover, some studies have been conducted on investigating eye responses on emotional stimuli [5,24,25].

In this study, we investigate whether eye activity holds discriminative power for recognising the emotional state, i.e. positive and negative emotions. We extract features for eye movements, pupil dilation and pupil invisibility to the eye tracker using a Tobii X120 eye tracker. We examine the performance of these features using artificial intelligence techniques for classification on an emotional stimulation experiment of 30 young adult and 30 mature adult subjects. Beside the machine learning techniques, we also analyse the statistical significance of the features between the two emotion and age groups. To separate memory from emotional effect on eye activities, we also compare eye activity of participants who have seen the positive clip of participants who have not.

The remainder of the paper is structured as follows. Section 2 reviews related background literature on using pupil features in affective studies. Section 3 describes the methodology, including the data collection, feature extraction and both statistical and classification methods. Section 4 presents the results. The conclusions are drawn in Section 5.

2   Background

Psychology research on pupil dilation has shown that not only light contributes to the pupil's response, but also memory load, cognitive difficulty, pain and emotional state [4]. Only few studies have been conducted for investigating eyes' response to emotional stimuli [5,13,24,25]. In the [13] study, it has been found that extreme dilation occurs for interesting or pleasing stimuli images. In [24,25], while their pupil dilation was recorded, subjects listened to auditory stimuli of three emotional categories: neutral, negative and positive. In [25], it has been found that pupil size was significantly larger during both emotionally negative and positive stimuli than during neutral stimuli. Using more controlled stimuli in [24], results showed that the pupil size was significantly larger during negative highly arousing stimuli than during moderately arousing positive stimuli. Another study [5] monitored pupil diameter during picture viewing to assess effects, showing that pupil size was larger when viewing pleasant and unpleasant emotional pictures.

One of the main issues with pupil variation measurement is the elimination of variation of pupillary light reflex, including materials with colour, luminance, and contrast. A review and suggestions of pupillary studies to eliminate luminance effects is presented in [11]. Avoiding visual stimuli variations, [24] and [25] used auditory stimuli. Another study used several statistical normalisation methods to minimise the variations in luminance, such as averaging pupil size and principal component analysis (PCA) [22]. Another way to fix this problem is to design a control paradigm that is identical for each subject to reduce variability [22]. Associating pupil measurement with other psychophysical measures, such as skin conductance, heart rate and brain signals could also be used to validate the reason of pupillary response [18]. In this study, we use both statistical normalisation methods and eliminate luminance effects (as far as possible).
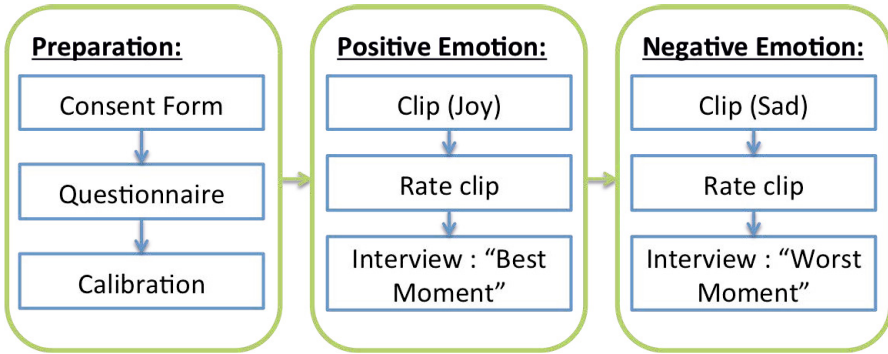
Fig. 1. Emotion eliciting paradigm and data collection process

However, there has been little research on eye activity that accompanies emotional responses to stimuli [16,28]. A study comparing eye activity as a response to positive and negative pictures, found greater eye blink reflex and corrugator muscle activity when viewing negative pictures [28]. In the [16] study, a decrease on eye blink and corrugator activity was found while suppressing negative emotions. It has been suggested that eye activity can help predict subjective emotional experience [21]. Therefore, in the current study, not only pupil dilation, but also the eye activity will be investigated to recognise the emotional state.

To the best of our knowledge, using an eye tracker sensor to analyse eye activity, pupil dilation and invisibility to the eye tracker is novel for the task of automatic emotion recognition. In this paper, we use a Tobii X120 eye tracker to extract eye activity features in order to analyse the differences between negative and positive emotions showing in eye behaviour in an emotion eliciting experiment. Beside analysing the general differences in eye movement, we specifically investigate the influence of age in the classification of emotions by comparing two age groups. After extracting features from each frame, we use a hybrid classifier where we build a Gaussian Mixture Model (GMM) for each subject's emotion, then feed it to Support Vector Machines (SVM) for classification. We also analyse the eye activity features statistically to identify how eye movement differs based on emotional state.

3   Method

A basic framework is designed to elicit positive and negative emotions only, using two video clips and two interview questions. The paradigm of our data collection is shown in Figure 1.

3.1   Emotion Elicitation Paradigm

Induced Emotions. Video clips have proven to be useful in inducing emotions [12] and been used for several emotional studies [30,17]. A universal list of emotional clips
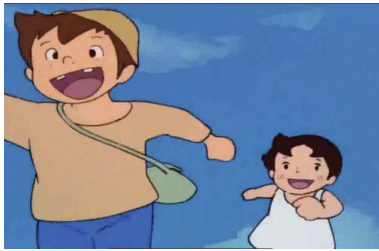
(a) "Heidi" (joy) clip   (b) 'Nobody's Boy: Remi" (sad) clip

Fig. 2. Sample video frames from the most highly rated video clips demonstrating positive (joy) and negative (sad) emotions

is available [12]; however; these clips are in English from 'Western society' movies. Considering cultural and language differences between the Western countries and Arab countries, it is possible that some of the validated clips, even when dubbed, will not obtain similar results. Moreover, using Arabic subtitles in those clips was not an option, since the measurement of the eye activity will be jeopardised. Given the unique culture of Saudi Arabia, where the study was conducted, and to ensure acceptance of all participants, an initial pool of 6 clips inducing positive and negative emotions was selected from classic cartoon animation series dubbed in Arabic. A basic survey to rate the emotion induction from those 6 clips was conducted on 20 volunteers. Those volunteers were not included in the latter eye activity data collection. The most highly rated video clips demonstrating positive (joy) and negative (sad) emotions were selected, namely: "Heidi" and "Nobody's Boy: Remi", respectively (see Figure 2). The two selected clips had almost similar duration ($\sim 2.5min$). The positive (joy) emotion clip shows a scene of rich depiction of nature and landscape where Heidi breathed fresh mountain air, felt the warmth of the sun on her skin, and happily met goatherd Peter. The negative (sad) emotion clip shows a scene of Remi learning the terrible truth of his beloved master Vitalis's death.

Spontaneous Emotions. Beside inducing emotions, watching video clips served as a preparation of the participant's mood for the subsequent spontaneous emotion recording part. To gain spontaneous emotions, participants were interviewed about emotional events in their life. That is, after watching the positive emotion clip, the participants were asked about the best moment in their life. For negative emotion, after watching the negative emotion clip, the participants were asked about the worst moment in their life.

3.2 Participants

The data collection procedure recruited 71 native Arabic speakers from a convenience sample (65 females, 6 males). The participants' age ranged from 18 to 41 years ($\mu = 25.6, \sigma = 4.8$). As regular participants' mood and mental state are important in the
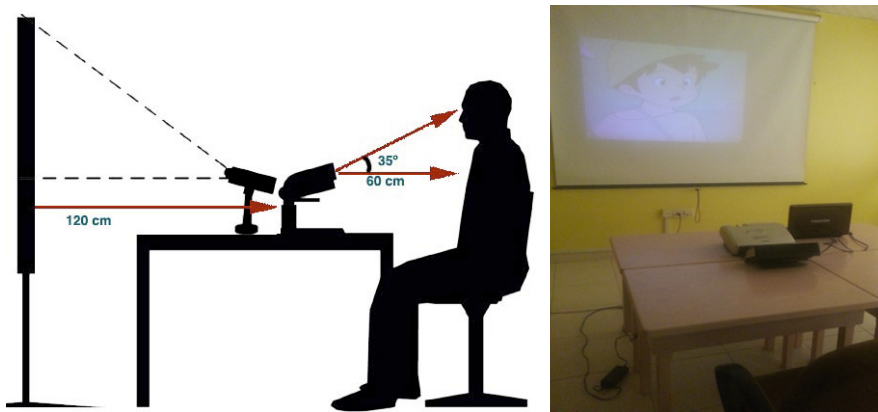
Fig. 3. Recording setup and environment

study, participants were asked about any current or history of mental conditions and about their usual mood: None of the participants had a mental disorder, 72% of the participants reported they are in a neutral mood, 7% always sad, and 22% always happy.

In this experiment, only 60 native female Arabic speakers were selected from the total recruited sample (30 young adults, and 30 mature adults), to insure age balance and reduce gender difference variability. The young adult participants' age ranged from 18 to 24 years ($\mu = 21.6, \sigma = 1.12$), while the mature adults' age ranged from 25 to 41 years ($\mu = 29.5, \sigma = 3.8$). Out of the selected subjects, 65% of participants had normal vision, the rest were using either glasses or lenses for correction.

3.3   Hardware and Recording Environment Settings

We used a Tobii X120 eye-tracker, attached to a Toshiba Satellite L655 laptop. We used a PowerLite 1880 XGA Epson projector screen as an extended monitor to the laptop, to ensure that the participant looked at similar coordinates while watching the clips and while talking to the interviewer. While the participant watch the clips, the interviewer leaves the room to reduce distraction and to allow the participant to freely watch the clips. The interviewer enters the room for the interview question and locates themselves in the middle of the projector screen. The screen resolution and distance from the projector screen and the eye tracker was fixed in all sessions. Although we had limited control over light in the recording room, we normalised the extracted features for each segment of each participant to reduce the light variability coming from the video clips themselves and the room light (see Figure 3).

3.4   Procedure

Consent and also a general demographic questionnaire asking about age, cultural heritage, physical and mental health, etc. were obtained prior to enrolling subjects in the study. Subjects were briefed about the study and were tested individually. Before the

beginning of the experiment, the subjects were instructed that they were going to watch the clips (without mentioning the specific type of emotion state) and told that the all film clips will be in Arabic. They were asked to watch the film as they would normally do at home and were told that there would be some questions to answer afterwards about the film clip and about their feelings. The eye movements of the subject were calibrated using a five-point calibration. This calibration was checked and recorded and upon successful calibration the experiment was started. Subjects were shown the instruction element screen asking them to clear their mind of all thoughts and then the clip began. After each clip, a post-questionnaire was done asking whether the subject had seen the clip previously, to investigate pupillary response due to memory activity. Moreover, to validate the induced emotion from the clips, participants were asked to rate the emotional effect of each clip in 11-points scale as: 'none (score: 0) to 'extremely (score: 10). Once the recording is over, subjects were thanked and no incentives were given.

Normalisation. To eliminate variability of pupil response not caused by emotional response, several aspects were considered:

- The design of the collection paradigm was controlled to be identical for each subject to reduce variability. The same clips were shown to each subject and also same interview questions asked.
- The screen resolution and distance from the projector screen and the eye tracker was fixed in all sessions. Further, recordings were done in the same room with similar daylight conditions.
- A projector screen was selected over a monitor screen, as once the interview starts, the interviewer locates themselves in the middle of the projector screen, to be in a similar position as the clips and to the eye-tracker.
- Moreover, percentile statistical normalisation methods have been applied to the extracted features from each subject to reduce within subjects variability as described bellow.

Preparation for Analysis. For each subject recording, clip watching and interview questions tasks of both positive and negative emotions were segmented using Tobii Studio (version 2.1). Having a total of 4 segments per subject, we extract raw features from each segment. To get clean data, we exclude frames where the eyes were absent to the eye tracker. The absence of the eyes is determined by the confidence level of the eye tracker of each eye (range 1 to 4). We only select frames where both confidence level of both eyes equal to 4.

3.5   Feature Extraction

Low-level Features. Excluding frames where the eyes were not detected by the eye tracker, we calculated 9 features per frame (30 frames per second) of raw data extracted from the Tobii eye tracker, as follow:

- Distance between eye gaze points position from one frame to the next for each eye, and its speed (Δ) and acceleration (Δ,Δ) were calculated to measure the changes

in eye gaze points; the longer the distance the faster the eye gaze change (2×3 features).

– Difference between the distance from left eye and the eye tracker and the distance from right eye and the eye tracker were calculated, to approximately measure head rotation (1 features).

– Normalised pupil size for each eye, to measure emotional arousal (2×1 features).

Statistical Features. Over the low-level features mentioned above, we calculated 147 statistical functional features to measure the pattern of eye activity, those features are:

– The mean, standard deviation (std), variance (var), maximum, minimum and range for all low-level features mentioned above (6×9).

– Even though, blink features are not available in Tobii X120, we measured the absence of the pupil in the video frames. Absence of left pupil only indicates left head rotation, and vice versa. Absence of both pupils could represent blinks, occluded eyes or head rotation being out of eye tracker range such as extreme looking up/down or left/right. We measure the mean, standard deviation (std), and variance (var) of the absence of left, right and both pupils (3×3)

– We also calculate several statistical features such as maximum, minimum, range and average of the duration, as well as its rate to total duration and count of occurrence of:

  • fast and slow changes of eye gaze for each eye ($6 \times 2 \times 2eyes$).
  • left, and right head rotation, calculated from the differences of the distance from both eyes and the eye tracker (6×3).
  • Large and small pupil size for each eye ($6 \times 2 \times 2eyes$).
 • The absence of left, right and both eyes (6×3)

The above duration features are detected when the normalized feature in question is higher than a threshold. The threshold is the average of the normalized feature in question plus the standard deviation of that feature for each segment.

3.6   Statistical Test

In order to characterise the eye activity patterns, the extracted statistical functionals from positive and negative emotions were compared. A two-tailed T-test was used for this purpose. In our case, the two-tailed T-tests for two samples were obtained assuming unequal variances with significance $p = 0.05$. The state of the T-test were calculated to identify the direction of effect.

3.7   Classification and Evaluation

For the low-level features, a Gaussian Mixture Model (GMM) with 8 mixture components was created for each segment for each participant. In this context, the GMM serves as dimensionality reduction, as well as a hybrid classification method [2]. The Hidden Markov Model Toolkit (HTK) was used to implement a HMM using one state to train the GMM models. In this work, diagonal covariance matrix was used, and the number
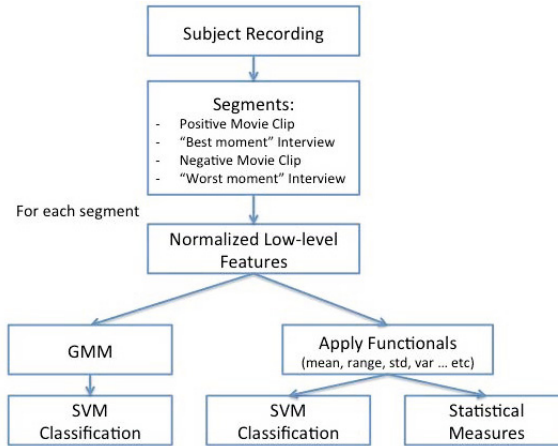
Fig. 4. Structure and Steps of the System

of mixtures was chosen empirically then fixed to ensure consistency in the comparison. This approach was beneficial to get the same number of values of the extracted features that to be fed to the Support Vector Machine (SVM) regardless of the duration of the participant's segment. The means, variance and weight of the 8 mixtures of GMM were used as a supervector that have been fed to the SVM classifier for each subject.

To test the effect of the eye activity patterns in each emotion (positive and negative) on the classification results, SVM was used on the statistical functional mentioned earlier. Comparing the low-level features modelling with statistical functional features classification is also beneficial to identify best modeling method for the task.

The segments for all participants were classified in a binary subject-independent scenario (i.e. positive / negative) using Support Vector Machine (SVM), which can be considered as a state-of-the-art classifier for some applications since it provides good generalisation properties. In order to increase the accuracy of the SVM, the cost and gamma parameters need to be optimized. In this paper, we used LibSVM [6] to implement the classifier, with a wide range of grid search for the best parameters. To mitigate the effect of the limited amount of data, a leave-one-subject-out cross-validation was used, without any overlap between training and testing data.

The main objective was to correctly classify the segment of each subject as positive or negative based on the eye activity patterns. The performance of a system can be calculated using several statistical methods, such as recall or precision. In this paper, the average recall (AR) was computed. Figure 4 summarize the general structure and the steps of our system.

3.8   Feature Selection

In order to maximise the recognition rate, manual and automatic feature selection were experimented with on the statistical features. Manual selection is based on the statistical tests mentioned above, as we manually select features that passes the T-test from

mature, young and all participants groups. For automatic feature selection, principal component analysis (PCA) [26] is a dimensionality reduction method, where high-dimensional original observations are projected to lower dimensions called principal components, which maximize the variance. As a result, the first principal component has the largest possible variance and so on for the next principal components. In this study, we perform a PCA on the statistical features, then used only the first 20% of the principal components for the classification, to be comparable (in number of features) with the manual feature selection.

4   Results

### 4.1   Initial Observations

Due to ethic restrictions in King Saud University regarding video-recording of participants, observations have been made only by the interviewer at the time of the interview and were not recorded. Regarding negative emotions, while watching the clip, 39% of participants rated the clip to have strong affect (more than 8 out of 10), though only almost 1% cried over the clip. On the other hand, while answering negative emotion interview question, 70% of the participants cried (including one male participant). Since the negative clip shows a death scene, almost 85% participants talked about their negative emotion during losing a loved person in their life. Other topics included injustice, failure, and conflict with a close person. Those late findings, indicate that watching the video clips prepared the participant mood for the spontaneous emotions in the interview. Since the number of male participants was not enough to make any reliable gender comparisons, more data needs to be collected. For the positive emotion, while watching the movie clip, 53% of participants rated the clip to have strong affect (more than 8 out of 10). On the other hand, while answering the positive emotion interview question, only 0.7% of the participants cried while expressing their joy (none of which were males). Our observations, indicate that unlike happiness crying, sadness crying was associated with eye contact avoidance.

As mentioned earlier, subjects were asked if they have seen the clip before the experiment. For the joy clip, 57% of the participants have seen the clip before. On the other hand, only 10% have seen the sad clip. Therefor, and to examine how the memory might effect the eyes activity, participants who have seen the joy clip were compared with participants who have not seen the clip, as can be seen later on. Figure 5 shows the number of participants who have seen the joy clip of each group.

Moreover, participants were asked to rate how much each clip had an affect on their feelings in scale from 0-10, given that 10 is a high affect and 0 has no affect. The joy clip got an average of 8 points as positive affect, and the sad clip got an average of 7 points as negative affect (see Figure 6).

4.2   Memory Effect on Eye Activity

In order to find out how memory could affect the eyes activity, participants were asked if they have seen each clip before. As mentioned earlier, 57% of the participants have
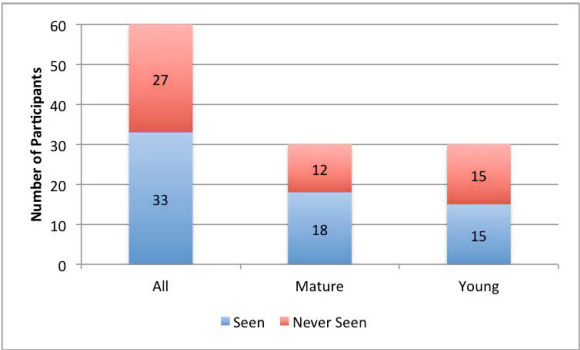
Fig. 5. Average rating of 'Joy' and 'Sad' clips of effectiveness in Inducing Emotions
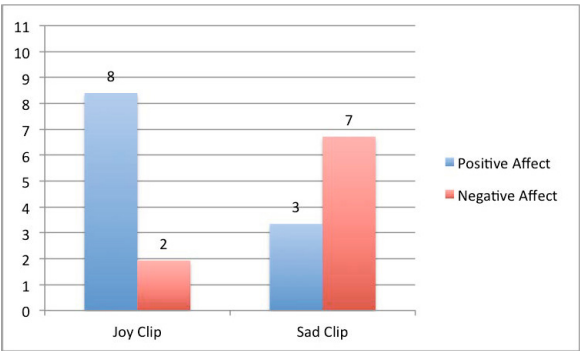
Fig. 6. Average rating of 'Joy' and 'Sad' clips of effectiveness in inducing emotions

seen the joy clip before, while only 10% have seen the sad clip. Since the number of participants who have seen the sad clip is very small, a comparison using the sad clip is not feasible. On the other hand, number of participants who have seen the joy clip is comparable to the ones who have not seen it. We investigated the differences between the two groups statistically and using artificial intelligent techniques.

Statistical Analysis. While comparing statistical features between participants who have seen the joy clip from participants who have not, only few features passed the t-test (see Table 1). We expected to get at least a slight difference in pupil size with memory provocation, as reviewed in [19]. However, it was not the case. The lake of significance might be due to the period since the clip have been seen. As the joy clip is extracted from classical animation series that was first aired in Saudi Arabia in 1983, and continued to air every two years till 1998 [8]. On the other hand, the only features that had significant differences was the speed and acceleration of changing gaze direction, being slower for participants who have seen the clip. This finding could support that memory for a

Table 1. Significant T-tests results of eye activity features Comparing Seen (S) vs. Unseen (N) Joy Clip for All Participants

Emotion	Feature	Direction	P-value
Joy Clip	Speed range of left eye gaze	S < N	0.050
	Speed range of right eye gaze	S < N	0.050
	Acceleration average of left eye gaze	S < N	0.040
	Acceleration average of right eye gaze	S < N	0.030

scene could have longer fixation duration and shorter saccade length, as this is still a controversial study, as reviewed in [27], more investigation is needed.

Classification Results. Classifying participants who have seen the clip against participants who have not, the average recall results were almost chance level (Table 2). Low-level features and functional features classification, as well as feature selection method are investigated to for any eye activity pattern differences from participants who have seen the clip against the ones who have not. The classification results supports our statistical analysis as there are no differences found by the classifiers. That finding might be due that the participants have seen the clip very long time ago, and they no longer remember the clip.

Table 2. Classification Average Recall Results in(%) for Participants Who Have Seen vs. Who Have Not Seen the Joy Clip

	Joy Clip		
Features	All Participants	Mature Adults	Young Adults
Low-level Features	48.7	37.5	41.7
All Statistical Features (146)	58.3	56.8	53.6
PCA Selection of Statistical Features (20%)	52.5	55.0	56.2

4.3   Emotional Effect on Eye Activities

In our paradigm, positive and negative emotions were elicited in two ways: using movie clips that induce emotions and by spontaneously answering questions about emotional events. In order to investigate the effect of elicited emotions on eye activity, we compared eye movements while experiencing positive and negative emotion statistically as well as artificial classification as follows.

Statistical Analysis. Regarding induced emotions from watching video clips, some statistical features were significant as shown in Table 3. The maximum and range of left and right pupils' size were larger in induced negative emotions. This finding is in-line with [24], where it has been found an increase in pupil size while listening to sound containing negative emotion. Even though blinking rate or duration could not be extracted using the eye tracker, we extracted the absence of the pupils in the video frames. Absence of both pupils could represent blinks, occluded eyes or head rotation being out of eye tracker range such as extreme looking up / down or left / right. Calculating the

Table 3. Significant T-test results of eye activity features comparing positive (P) vs. negative (N) emotions for all participants

Emotion	Feature	Direction	P-value
Induced	Maximum left pupil size	P < N	0.027
	Range of left pupil size	P < N	0.010
	Maximum right pupil size	P < N	0.026
	Range of right pupil size	P < N	0.012
	Average absence of both pupils	P < N	0.005
	Absence duration rate of both pupils	P < N	0.005
	Absence duration range of both pupils	P < N	0.038
	Maximum absence duration of both pupils	P < N	0.038
Spontaneous	Maximum duration of left head rotation	P < N	0.053
	Maximum duration of large left pupil size	P < N	0.014
	Maximum duration of large right pupil size	P < N	0.002
	Maximum absence duration of left pupil	P < N	0.035
	Absence duration range of left pupil	P < N	0.047
	Maximum absence duration of right pupil	P < N	0.004
	Absence duration range of right pupil	P < N	0.015
	Maximum absence duration of both pupils	P < N	0.017
	Absence duration range of both pupils	P < N	0.027

differences in the absence of the pupils for both positive and negative emotions could be used as an indicator of emotions. For example, sadness is characterised by downward eyebrows and head [10]. In our study, the statistical functional of absence duration of both pupils was significantly longer in negative emotions, which might be a sign of extreme head rotation, blinks, or occluded eyes. This finding is supported by the general characteristic of expressing sadness [10].

Table 3 shows some of the significant statistical features while expressing positive and negative spontaneous emotion in the interview. In this study, we found that the maximum duration of left head rotation (measured by the different in the distance between the eyes and the eye tracker) is significantly longer while expressing spontaneous negative emotions. That could be explained by avoiding eye contact, as there is evidence that during discussing personal topics and during speech hesitation there is less eye contact [3]. Regarding pupil size, we found that the maximum duration of large pupil size for both left and right eye were significantly larger in spontaneous negative emotions, which is inline with our finding in induced emotions. As in induced emotions, the absence duration of left, right and both pupils are measured. Range and maximum absence duration of left, right and both pupils were longer while expressing negative emotions, indicating extreme head movements and occluded eyes. Giving that more than %70 of the participants cried while being interviewed in the negative emotion questions, the absence of both pupils are explained by the occluded eyes.

Classification Results. Automatic classification of positive and negative emotions from eye activity features were performed and the results are shown in Table 4. In general, the recognition rate of emotion types from eye activities was reasonable, giving 66% on average. This indicate that eye activities could be used as a complementary cue in identifying emotional states. Although both induced and spontaneous emotions have relatively reasonable recognition rates (65.3% and 66.7% on average, respectively), generally while expressing spontaneous emotions the average recognition

Table 4. Average recall results in (%) for positive vs. negative emotion classification

Features	Induced Emotions			Spontaneous Emotions		
	All Participants	Mature-adults	Young-adults	All Participants	Mature-adults	Young-adults
Low-level Features	60.8	66.4	62.3	65.1	68.5	68.4
Manual Selection of Statistical Features (passed T-test)	68.3	68.3	75	65.8	68.3	71.6
PCA automatic selection of Statistical Features (20%)	60.1	60	66.6	64.2	63	65

rate of positive and negative emotions is slightly higher than induced emotions. This finding indicates that spontaneous emotions might have stronger eye activity patterns than induced emotions.

Unlike statistical features, low-level features modelled only eye activity excluding pupils invisibility to the eye tracker, where the pupils of either eyes were not detected. Regardless of the high dimensionality, modelling functional features performed approximately equivalent to low-level features, if not better. Current findings indicate that statistical modeling loses less information compared to the low-level modeling. Comparing manual (T-test) and automatic (PCA) feature selection of the functionals, the manual selection based on the T-tests results was better than using the automatic selection with PCA.

Previous studies have investigated age effects on emotion intensity, finding a reduction in intensity between young adult and middle aged groups [20]. In our study, mature adults sustained almost equivalent recognition rates for both induced and spontaneous emotions. On the other hand, young adults' recognition rate for both induced and spontaneous emotion was significantly higher than mature adults. The sustained recognition rate in the spontaneous emotion for mature adults might be explained by the adaptation, i.e. older people have been exposed to more emotional incidents and, thus, been adapted [20]. Regarding young adults, the finding is not surprising as in [20] it was expected to have better recognition rate in young adults for both induced and spontaneous emotions.

5   Conclusions

Automatic detection of human emotional states has had much interest lately for its applications not only in HCI, but also for its applications in psychological studies. Given the rich source of the eye activities regarding the emotional state, we investigated whether eyes activities, including movements, and pupil dilation or invisibility to the eye tracker hold discriminative power for detecting emotional states.

A framework to elicit positive and negative emotions was implemented using two video clips for induced emotions and using interview for spontaneous emotions. We collected eye movements from 71 participants, were only 60 female participants were selected for the analysis to reduce gender variability. We extracted several low-level features as well as applying statistical functionals over them. We analyzed statistical functional features statistically as well as using artificial technique classifications.

In general, the classification results using eye activities were reasonable, giving 66% on average, proving that eye activities holds effective cues in identifying emotional states. Generally while expressing spontaneous emotions, the recognition rate of positive and negative emotions is slightly higher than induced emotions. This finding indicates that

spontaneous emotions might have stronger eye activity patterns than induced emotions. Comparing low-level eye activity features with statistical features, we found that statistical features modeling is as good in capturing information compared to the low-level modelling, therefore, performed better in recognising emotional states. We also investigate manual and automatic feature selection, the different between classification results from both methods were statistically significant, we conclude that manual selection based on T-test performs better.

We found that pupil dilation size and duration increase while expressing negative emotions. We also found less eye contact, explained by head rotation. Calculating the absence of pupils in the video frames indicates extreme head rotation away from the camera and occluded eyes was longer while expressing negative emotions.

Investigating the age effect on expressing emotions, we compared young and mature adults. While mature adults sustained almost equivalent recognition rates for both induced and spontaneous emotions, the young adults' recognition rates for induced and spontaneous emotion were significantly higher than mature adults. Moreover, the sustained recognition rate in mature adults with induced and spontaneous emotions, might be due to adaption.

To separate memory from emotional effect on eye activities, we compared eye activities from participants who have seen the joy clip from participants who have not. The classification results were almost chance level, as there were no differences from the two groups. Statistical analysis showed no significant different in all features but slower speed of change gaze direction for participants who have seen the clip. As this finding is still controversial in the psychology literature, more investigation is needed using more participants.

We conclude that eye activity including eye movement, pupil dilation and invisibility to the eye tracker could be a complementary cues for the automatic recognition of human emotional states.

6   Limitation and Future Work

A known limitation in this study is the lack of an official rated list of eliciting emotion Arabic clips. A separate study investigating affective Arabic clips is needed for such cultural emotional studies. Future work in this current collected database will analyse speech and investigate its correlation with eye activities. The current database collection was limited to voice recording and eye tracking, future collection will expand not only on the number of participants, but also on using more sensing devices such as skin conductivity, temperature, heart rate, and Brain-Computer Interfaces (BCI) including measuring eye blinking duration and rate.

References

1. Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Breakspear, M., Parker, G.: From Joyous to Clinically Depressed: Mood Detection Using Spontaneous Speech. In: Proc. FLAIRS-25 (2012)

2. Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Breakspear, M., Parker, G.: A Comparative Study of Different Classifiers for Detecting Depression from Spontaneous Speech. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (1), pp. 26–31 (2013)

3. Argyle, M., Dean, J.: Eye-contact, distance and affiliation. Sociometry, 289–304 (1965)

4. Beatty, J.: Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychological Bulletin 91(2), 276 (1982)

5. Bradley, M.M., Miccoli, L., Escrig, M.A., Lang, P.J.: The pupil as a measure of emotional arousal and autonomic activation. Psychophysiology 45(4), 602–607 (2008)

6. Chang, C.C., Lin, C.J.: Libsvm: a library for svm, pp. 1–30 (2001), www.csic.ntu.edu.tw/rcjlin/papers/lib.svm (March 4, 2006)

7. Craig, S., Graesser, A., Sullins, J., Gholson, B.: Affect and learning: an exploratory look into the role of affect in learning with autotutor. Journal of Educational Media 29(3), 241–250 (2004)

8. Ministry of Culture, A, Information.: Saudi arabia channel one (2013), http://www.sauditv1.tv/

9. Duchowski, A.T.: A breadth-first survey of eye-tracking applications. Behavior Research Methods 34(4), 455–470 (2002)

10. Frijda, N.H.: The emotions. Cambridge University Press (1987)

11. Goldwater, B.C.: Psychological significance of pupillary movements. Psychological Bulletin 77(5), 340 (1972)

12. Gross, J.J., Levenson, R.W.: Emotion elicitation using films. Cognition & Emotion 9(1), 87–108 (1995)

13. Hess, E.H.: Pupillometrics: A method of studying mental, emotional and sensory processes. In: Handbook of Psychophysiology, pp. 491–531 (1972)

14. Hobson, R.P., Ouston, J., Lee, A., et al.: Emotion recognition in autism: Coordinating faces and voices. Psychological Medicine 18(4), 911–923 (1988)

15. Horng, W.B., Chen, C.Y., Chang, Y., Fan, C.H.: Driver fatigue detection based on eye tracking and dynamk, template matching. In: 2004 IEEE International Conference on Networking, Sensing and Control, pp. 7–12. IEEE (2004)

16. Jackson, D.C., Mueller, C.J., Dolski, I., Dalton, K.M., Nitschke, J.B., Urry, H.L., Rosenkranz, M.A., Ryff, C.D., Singer, B.H., Davidson, R.J.: Now you feel it, now you don't frontal brain electrical asymmetry and individual differences in emotion regulation. Psychological Science 14(6), 612–617 (2003)

17. Jerritta, S., Murugappan, M., Nagarajan, R., Wan, K.: Physiological signals based human emotion Recognition: a review, pp. 410–415. IEEE (2011)

18. Kahneman, D., Tursky, B., Shapiro, D., Crider, A.: Pupillary, heart rate, and skin resistance changes during a mental task. Journal of Experimental Psychology 79(1p1), 164 (1969)

19. Laeng, B., Sirois, S., Gredebck, G.: Pupillometry: A window to the preconscious? Perspectives on Psychological Science 7(1), 18–27 (2012)

20. Larsen, R.J., Diener, E.: Affect intensity as an individual difference characteristic: A review. Journal of Research in Personality 21(1), 1–39 (1987)

21. Niemic, C.P., Warren, K.: Studies of emotion. A Theoretical and Empirical Review of Psychophysiological Studies of Emotion (Department of Clinical and Social Psychology). JUR Rochester 1(1), 15–19 (2002)

22. Oliveira, F.T., Aula, A., Russell, D.M.: Discriminating the relevance of web search results with measures of pupil size. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, pp. 2209–2212. ACM (2009)

23. Pantic, M., Pentland, A., Nijholt, A., Huang, T.: Human computing and machine understanding of human behavior: a survey. In: Proceedings of the 8th International Conference on Multimodal Interfaces, pp. 239–248. ACM (2006)

24. Partala, T., Jokiniemi, M., Surakka, V.: Pupillary responses to emotionally provocative stimuli. In: Proceedings of the 2000 Symposium on Eye Tracking Research & Applications, pp. 123–129. ACM (2000)

25. Partala, T., Surakka, V.: Pupil size variation as an indication of affective processing. International Journal of Human-Computer Studies 59(1), 185–198 (2003)

26. Pearson, K.: Liii. on lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2(11), 559–572 (1901)

27. Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychological Bulletin 124(3), 372–422 (1998)

28. Sutton, S.K., Davidson, R.J., Donzella, B., Irwin, W., Dottl, D.A.: Manipulating affective state using extended picture presentations. Psychophysiology 34(2), 217–226 (2007)

29. Tao, J., Tan, T.: Affective computing: A review. In: Tao, J., Tan, T., Picard, R.W. (eds.) ACII 2005. LNCS, vol. 3784, pp. 981–995. Springer, Heidelberg (2005)

30. Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(1), 39–58 (2009)

31. Zhou, F., Ji, Y., Jiao, R.J.: Affective and cognitive design for mass personalization: status and prospect. Journal of Intelligent Manufacturing, 1–23 (2012)