Evaluation of Signal Processing Methods for Attention Assessment in Visual Content Interaction

Georgia Elafoudi[1(✉)], Vladimir Stankovic[1], Lina Stankovic[1], Deepti Pappusetti[2], and Hari Kalva[2]

[1] Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK
{georgia.elafoudi,vladimir.stankovic,lina.stankovic}@strath.ac.uk
[2] Department of Computer and Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL, USA
{dpappuse,hari.kalva}@fau.edu

Abstract. Eye movements and changes in pupil dilation are known to provide information about viewer's attention and interaction with visual content. This paper evaluates different statistical and signal processing methods for autonomously analysing pupil dilation signals and extracting information about viewer's attention when perceiving visual information. In particular, using a commercial video-based eye tracker to estimate pupil dilation and gaze fixation, we demonstrate that wavelet-based signal processing provides an effective tool for pupil dilation analysis and discuss the effect that different image content has on pupil dilation and viewer's attention.

1 Introduction

Objectively assessing users' experience when interacting with visual content is gaining increased research interest due to its relevance for numerous applications ranging from video compression, Human Computer Interaction (HCI)-based decision support tools design, web-site design, and Internet visual searches, to those related to marketing, science and medicine. Eye trackers that use high-resolution, high-speed video cameras to record eye movements and corneal reflections are cost-effective tools for high-precision measurement of the size of the pupil, gaze locations and the time length of fixation. They consist of a video camera and infrared illuminators, positioned in front of an eye, used to track the movement of the eyes which are then mapped into real-world coordinates by camera calibration.

The gaze location and the time length of fixation obviously show which features of the image the user is looking at and can reveal, for example, which features attract "the eye", which features are missed, and identify the point of

We would like to thank all participants in the study. The work was supported by FP7 QoSTREAM project, http://www.qostream.org.

V. Murino et al. (Eds.): ICIAP 2015 Workshops, LNCS 9281, pp. 580–588, 2015.
DOI: 10.1007/978-3-319-23222-5_70

visual fixation. However, since purely looking at an image feature does not necessarily mean that the feature attracted attention and caused the desired cognitive reaction, relating pupil dilation to cognition is a promising research direction. Numerous studies in psychology have demonstrated that changes in pupil dilation consistently occurs during the cognition process, including reading, visual search, and problem solving, as well as valence, arousal, pain, etc. (see [1]-[10]).

Extracting useful information from raw pupil dilation signals is not a trivial task due to high measurement noise of the commercial camera-based eye trackers, distortion due to gaze angle [1], frequent eye blinking, effects of illumination changes, irregular time delays in pupil response to stimuli, as well as the fact that pupil reaction is caused by different factors, which are hard to separate. In this paper, we designed a set of simple experiments based on a commercially available eye tracker, and use them to evaluate several statistical and signal processing tools for extracting useful information from pupil dilation signal when a user is presented with a sequence of coloured images. We demonstrate that mean, peak and variance are insufficient to capture the dynamics of pupil dilation change and propose frequency-based and wavelet-based analysis, for defining and extracting features that can be further used for clustering or pattern matching.

2 Background

Pupils respond to different stimuli, including pain, emotional reaction, mental workload, and arousal with a very uneven reaction time delay and intensity that depends on the intensity and type of stimuli. The reaction delay can range from 0.1sec (mainly in the case of pain) to 2-7 seconds for emotional stimuli [1]. For example, [2] investigated reaction to sound stimuli, and observed that only after 400ms the pupil starts to sharply dilate, reaching a peak 2-3sec after the stimuli.

Beatty [3] concluded that pupil dilation is a good representation of the difficulty and amount of mental workload across tested subjects and cognitive tasks. For example, calculating 16×23 causes 10% larger pupil dilation than 7×8, or memorizing a 3-digit number and 7-digit number caused 0.1mm and 0.55mm of pupil dilation, respectively (see [1]).

Different measures, such as mean dilation, peak dilation, variance, as in [4], and response time, have been proposed to quantify pupil dilation as a response to cognitive tasks (see [1] and references therein). However, these measures are not very robust and often not informative enough.

The key challenge in extracting meaningful information from pupil dilation lies in distinguishing the exact cause of pupil reaction. The preprocessing task needs to remove distortion and camera noise as well as natural blinking and pupillary light reflex. Indeed, as a reaction to brightness change, pupils naturally dilate, which can significantly affect measurements. To mitigate this problem, in [5] and [6], principal component analysis (PCA) is used on the pupil dilation data. Another approach can be found in [7], where a Hilbert transform method was used in order to study cognitive overload and cognitive dissonance. Though the initial results show potential, they are not conclusive, according to the authors, requiring further studies.

[8] proposed a measure called *Index of cognitive activity* (ICA), that represents the average number of "abrupt" changes in pupil size per second. This was estimated using wavelet decomposition, a technique that proved capable of filtering out the change of brightness effect. Building on this work, Marshall [9] compared pupil dilation with other measures such as blink rates, fixation time, saccade distance and speed, during different tasks, such as driving a car and visual search, in order to identify the best combination of measures for assessing the cognitive state of a subject. This research has proposed a combination of seven "eye metrics", left and right index, left and right blink, left and right movement, and divergence.

All these methods are either limited in conclusions they can make due to a restricted extracted feature space, or require additional measurands. Despite the fact that the pupil dilation signal has been studied for long (see [1] and references therein), there are still many unknowns w.r.t exploiting pupil dilation analysis for assessing multimedia experience, and appropriate signal processing and machine learning tools are needed to make the information extraction process fast and automated, to open the door for real-time visual feedback design mechanisms.

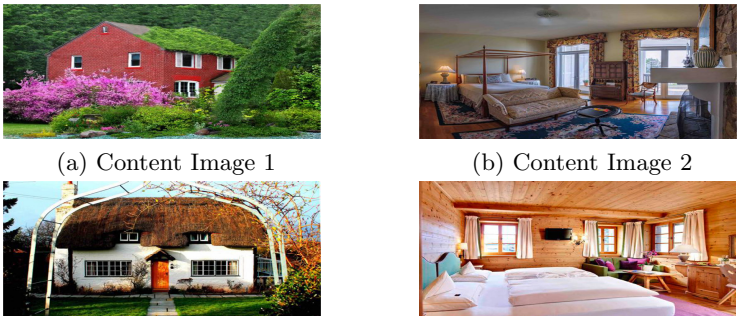
3 Methods

Experiments were designed to examine the gaze position and pupil responses to different, "neutral-content" images (indoors and outdoors) with and without searching for a specific target. The "neutral-content" images were selected as we want to examine pupil activity that is not a product of emotional triggering (pleasant/unpleasant), which is expected to create a more intense response. Additionally, we want to assess the effect of "busy" indoor images versus less busy outdoor images.

The eye tracker used during all our experiments is the Tobii X2-60 Eye Tracker, which provides pupil dilation and gaze fixation data with a 60Hz sampling rate. Data collection and stimuli presentation were obtained using Ogama.

Experimental Setup. Experiments were performed in a laboratory using moderate artificial light conditions, which were kept constant for the duration of all trials. Ten subjects who participated in the experiment ranged between 25 and 50 years old, both male and female, either with normal or corrected-to normal vision. The subjects were sitting in front of a screen with a resolution of 1920x1080 pixels, at a $\sim$ 70 cm distance (35° angle from the eye tracker). Calibration was performed using Ogama's calibration process, where a coloured dot was moving in the corners and centre of the screen and the subjects were asked to gaze at the dot. This process was performed before each trial.

Stimuli. The stimuli for this experiment were four high quality images that were acquired from Flickr under Creativity Commons Licences and are shown in Fig. 1. Two of the images represent the outside of residential properties and the other two are indoor bedroom images. Outdoor images did not resemble the architecture of the area where the test subjects reside. Between each image, and



(a) Content Image 1 (b) Content Image 2

(c) Content Image 3 (d) Content Image 4

Fig. 1. Content images used as a stimuli during the experiments.[1]

at the beginning and end of the presentation, a whole grey-coloured blank image of the same size was used to separate each different stimuli. Each image (stimuli or grey) was shown for 10 sec.

For each subject, the experiment comprised three trials, which took place at the same time and place, with a gap of less than a minute between the two trials. During the first trial (Trial 1), the subjects were shown the stimuli presentation, and were asked to watch the presentation with no further instructions. After the end of the trial, the same stimuli was shown for the second time (Trial 2), with no further instructions. During the final trial (Trial 3), all subjects were requested verbally to locate the flower(s) in the images at the beginning of the trial, without doing any task when this occurred or without verbally suggesting that they had identified the target.

Preprocessing. Pupil dilation and gaze data were cleaned during the preprocessing by removing all eye blink artifacts, as in [4,6]. All missing data were replaced using data interpolation from both right and left pupil data using linear interpolation. We perform three types of signal processing analysis on the processed data: (1) statistical analysis using dilation mean, variance, and peak; (2) frequency analysis; and (3) wavelet-based analysis.

Harmonic Analysis is performed using Welch Method [12] which is a commonly used method for estimating the *power spectral density* (PSD) of a signal in the presence of noise. It splits the data into overlapping segments, computes modified periodograms of the overlapping segments and then averages them in order to estimate the PSD and mitigate effects of random noise. In the applied procedure, the signal is segmented into eight sections of equal length, each with 50% overlap. All remaining signal parts that cannot be included into these eight

[1] Image 1 was taken from Billy Wilson under CC BY-NC 2.0, Image 2 and 4 from Kay Gaensler and Marketing Deluxe, respectively, under CC BY-NC-SA 2.0, and Image 3 from Les Haines under CC BY 2.0.

segments are discarded. Each segment is windowed with a Hamming window of the same length as the segment.

Wavelet-based Analysis. After interpolation and prior to wavelet-based analysis, to mitigate the effect of random measurement noise, the pupil dilation signal was filtered by a 5th order low-pass Butterworth filter with a cutoff frequency of f_c =4Hz, which was selected as in [4] and [6], since the pupil servomechanism's break frequency is roughly 2Hz (see [6] and references therein).

After filtering, we performed *Discrete Wavelet Transform* (DWT) decomposition of the signal and wavelet denoising using soft thresholding [11]. This is done by passing the signal through a low-pass and a high-pass filter, and then down-sampling the filtered signals in order to remove the over-completeness of the transform coefficients. Due to the properties of DWT, the energy of the transformed signal is concentrated in only few DWT coefficients that have high magnitudes, and the energy of the noise is spread across a large number of DWT coefficients that have low magnitudes. Wavelet Denoising by Soft Thresholding [11] can be applied to remove the remaining noise in the 0-4Hz band, by minimizing mean square error (MSE) of the reconstructed signal compared to the original signal under the constraint that with high probability the reconstruction is at least as smooth as the original. This allows for the removal of undesirable noise ripples or oscillations that would not be removed with a simple MSE minimization. The idea of wavelet denoising by soft thresholding is to first decompose the noisy signal into N levels using a pyramidal wavelet filter, and then apply thresholding on the wavelet coefficients coordinate-wise with a specially selected threshold. All DWT coefficients whose absolute value is less than the predefined threshold are set to zeros and all remaining coefficients will have magnitude reduced by the applied threshold. In the proposed method, we use *Minimax thresholding* and estimate the level of noise based on the first level coefficients. Finally, the inverse transform is applied to recover the original signal. We note that a similar wavelet denoising procedure was used in [8] to evaluate the level of cognitive activity based on pupil dilation.

4 Results and Discussion

We separated the filtered pupil dilation signal into image segments and then concatenated all content image segments and all grey image segments, forming in this way two signals: a signal carrying four content images and a signal carrying grey images. Fig. 2 shows the filtered pupil dilation signal for Subject 5 and 10 during all three trials for the content images. Each vertical line represents the temporal transition between images. These subjects were selected as a representative example of all subjects. The pupil dilation range was around 1mm for both subjects and it dynamically changed during the experiments. From these graphs it is apparent that the dilation on average is higher during Trial 3, compared to the other two trials.

In Fig. 3 we show the gaze fixation for Subjects 5 and 10 during Trial 3, where the subjects were asked to locate "flower(s)" in the images. Figs. 3 (a)

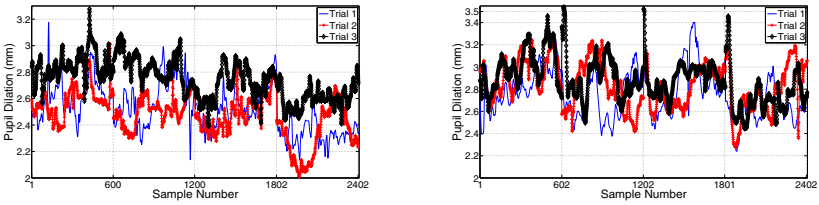


Fig. 2. Filtered Pupil Dilation for Subjects 5 and 10 respectively during all three trials

and (d) show Gaze Position (X,Y) versus Time, i.e., the sample number. Figs. 3 (b)-(c) and (e)-(f) illustrate the gaze position for both subjects for each content image. It is apparent by looking at the original images in Fig 1 that clusters of gaze points are located in the image areas with flowers. Indeed, for Image 2, in Fig. 3(c), we can observe that Subject 5 was able to target the living flowers. On the other hand, Subject 10 noticed many artificial flowers in the image. Note that there was no specification about the type of flower(s) the subjects should locate. Similarly, Images 1 and 3 have multiple flowers, hence the multiple clusters of gaze fixation points. In Image 4, both subjects were able to locate the vase on the table. The 3D graphs provide extra information of when the different targets were identified, spread out across the period the image(s) were shown.

Table 1 shows the Mean values of pupil dilation when viewing the Content Images, after filtering. From Table 1, we can see that predominately the highest values (bold) of mean pupil dilation were during Trial 3, which is expected.

In general though, mean, variance and number of peaks (not shown here due to space restrictions) rather irregularly change across the images and subjects, failing to capture signal transients. Thus, time averaging over the images do not, in this case, provide information that can be used to assess user's attention and experience, since the signal transition information is lost.

Harmonic Analysis. Next, we performed harmonic analysis by estimating the PSD of the signal in the range of interest (0-2Hz) using the Welch method described above. We separately estimated PSD for each grey image and each content image. The results for all content images are shown in Table 2. Frequency analysis shows the distribution of the power indicating in which frequency sub-band most of signal's energy is concentrated. It could potentially indicate increased mental activity, if small enough time windows are applied. This can be seen from Table 2 as on average Trial 3 shows increased power values per image, when compared with other trials. For example, Subject 5 in Image 1 shows 564.4W, which is higher compared to 477.6W and 468.5W for Trials 1 and 2, respectively. This pattern is similar to the case of mean values in Table 1 and generally power values are higher for Trial 3 when the subjects were asked to perform a target search task. This is mostly pronounced for Image 3 where all subjects show higher energy at Trial 3 (see bold values in Table 2).

Table 1. Mean of the Filtered Pupil Dilation Signal of the Content Images [in mm]. S stands for Subject and **Bold** represents the highest mean value per image

	Image1			Image 2			Image 3			Image 4		
S$	Trial1	Trial2	Trial3	Trial1	Trial2	Trial3	Trial1	Trial2	Trial3	Trial1	Trial2	Trial3
1	2.180	**2.193**	2.123	2.215	**2.236**	2.213	2.135	2.227	**2.315**	2.111	**2.258**	2.246
2	**2.576**	2.500	2.546	2.606	2.751	**2.795**	2.479	2.583	**2.745**	2.458	**2.481**	2.417
3	3.104	**3.241**	3.215	3.312	3.103	**3.429**	3.091	2.991	**3.097**	**3.038**	2.849	2.958
4	**3.109**	2.631	2.749	**3.157**	2.823	2.843	2.779	2.699	**2.808**	2.642	2.644	**2.703**
5	2.550	2.520	**2.771**	2.528	2.442	**2.765**	2.477	2.428	**2.640**	2.275	2.270	**2.530**
6	2.578	**2.689**	2.673	2.645	2.665	**2.833**	2.609	2.683	**2.781**	2.499	**2.608**	2.586
7	2.614	2.544	**2.657**	**3.023**	2.653	2.872	**2.680**	2.627	2.668	**2.830**	2.612	2.542
8	**3.402**	3.204	3.374	3.481	3.185	**3.500**	3.209	3.050	**3.350**	**3.265**	3.139	3.221
9	3.131	3.059	**3.134**	**3.186**	3.159	3.032	2.929	3.029	**3.163**	2.804	**2.829**	2.707
10	2.757	2.873	**2.925**	2.639	2.739	**2.828**	2.770	2.743	**2.828**	2.546	**2.690**	2.653

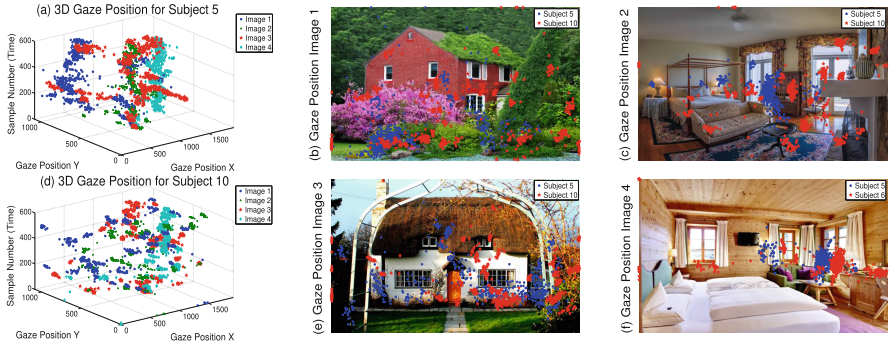


Fig. 3. Gaze Position for Subject 5 and 10 for all images (3D) during Trial 3, and individually per content image.

As frequency analysis loses time information and makes it difficult to conclude which time stimuli caused the reaction, we propose next wavelet-based analysis.

Wavelet-based Signal Processing. Fig. 4 shows the wavelet-based analysis for Subjects 5 and 10 for all trials. Similar results are obtained for other subjects. Horizontal axis again shows the sample number with vertical lines pointing to the image transition moments; vertical axis denotes the right eye pupil dilation in mm. We used Daubechies-4 wavelet filter which is one of the most popular orthogonal wavelet filter with fast wavelet transform. In contrast to the Fourier analysis, DWT can tradeoff frequency and time resolution allowing for detection of time interval when specific frequency component occurred. We used a 4-level wavelet decomposition, decomposing the signal into 4 frequency bands, to maintain high frequency resolution. The areas around the image transitions should be ignored as they are caused by signal stitching. We can clearly observe

Table 2. Signal Power in the 0-2Hz Band in [W] per content image. **Bold** represents the highest power per image.

	Image1			Image 2			Image 3			Image 4		
S	Trial1	Trial2	Trial3	Trial1	Trial2	Trial3	Trial1	Trial2	Trial3	Trial1	Trial2	Trial3
1	348.8	**355.4**	335.1	362.2	**367.7**	359.5	336.4	365.5	**392.2**	329.8	**373.2**	373.2
2	490.2	461.8	**506.9**	502.6	**557.4**	520.4	451.1	490.2	**491.1**	447.0	**451.9**	428.9
3	707.9	**772.9**	763.1	802.5	708.6	**863.1**	701.6	663.8	**703.2**	**677.3**	596.9	642.6
4	**710.8**	510.2	554.5	**731.0**	585.7	592.4	567.6	538.0	**580.0**	512.7	516.2	**538.9**
5	477.6	468.5	**564.4**	470.0	436.5	**559.0**	451.8	432.9	**512.6**	379.1	379.8	**470.9**
6	490.2	**530.6**	523.5	513.5	521.3	**585.3**	500.6	529.3	**567.0**	458.0	**498.6**	491.9
7	502.2	476.9	**523.2**	**672.6**	516.4	602.9	528.8	507.2	**531.7**	**620.3**	502.1	474.5
8	**855.8**	756.3	852.0	891.8	749.3	**910.0**	757.0	687.1	**823.2**	**782.6**	721.4	771.4
9	**726.8**	695.3	722.2	**746.2**	742.2	676.3	636.3	678.7	**736.3**	579.8	**590.6**	540.5
10	561.0	605.3	**624.4**	517.0	551.2	**588.4**	567.6	554.4	**587.6**	477.4	**539.6**	519.6

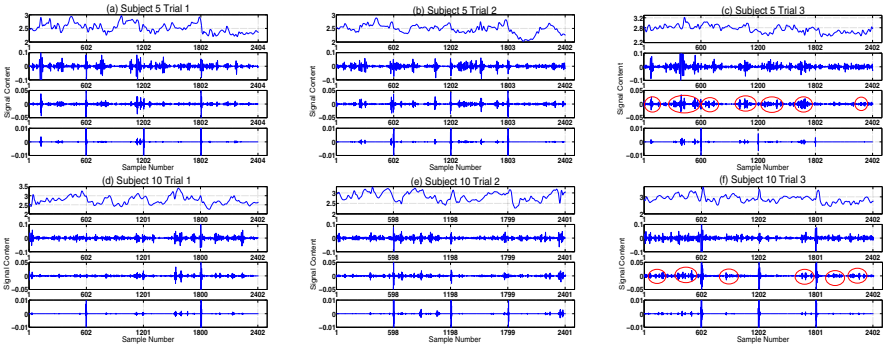


Fig. 4. Wavelet Analysis of Content Images for Subject 5 and 10 for all three trials.

increased activity in the band of interest. Trial 3 is characterised by significant pupil dilation activity in the beginning (until the task is solved, bearing in mind delayed reaction) and then reduction, while the other two trials have more evenly spread activity. Trial 1 has evidently more activity than Trial 2, since new content was presented in Trial 1. This activity has been marked for both subjects in Figs. 4(c) and (f) with red circles, where activity is more clear. In Trial 3, the pupil dilation activity indicates that Image 2 was most challenging, which is true since the flower position is not so obvious.

Conclusion. The paper discussed different signal processing methods for analyzing pupil dilation signals with applications to multimedia experience assessment. Our main goal was to review and test different methods, in order to evaluate their future use for feature definition and extraction for autonomous pattern matching and event detection. As in [8], our findings show that clearly

wavelets provide a clearer view of activity on pupil dilation and can be essentially used as a helping tool for extracting signatures from the pupil dilation signal in order to relate each segment to image information/task for automated pattern matching. In addition to the wavelets, mean, variance, and power spectral density using Welch method, can be used in order to provide a more accurate activity recognition process.

References

1. Wang, J.Y.-Y.: Pupil dilation and eye-tracking. In: Schulte-Mecklenbeck, M., Kuhberger, A., Ranyard, R. (eds.) A Handbook of Process Tracing Methods for Decision Research: A Critical Review and User's Guide. Psychology Press (2010)
2. Partala, T., Surakka, V.: Pupil size variation as an indication of affective processing. International Journal of Human-Computer Studies **59**(1), 185–198 (2003)
3. Beatty, J.: Task-Evoked Pupillary Responses, Processing 19, and the Structure of Processing Resources. Psychological Bulletin **91**(2), 276–292 (1982)
4. Privitera, C.M., Renninger, L.W., Carney, T., Klein, S., Aguilar, M.: The pupil dilation response to visual detection. In: Human Vision and Electronic Imaging (2008)
5. Oliveira, F.T.P., Aula, A., Russell, D.M.: Discriminating the relevance of web search results with measures of pupil size. In: Proc. 27th ACM CHI'2009 (2009)
6. Privitera, C.M., Renninger, L.W., Carney, T., Klein, S., Aguilar, M.: Pupil dilation during visual target detection. Journal of Vision **10**(10), 3 (2010)
7. Hossain, G., Yeasin, M.: Understanding effects of cognitive load from pupillary responses using hilbert analytic phase. In: CVPRW 2014, pp. 381–386 (2014)
8. Marshall, S.P.: Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity. Google Patents, US6090051 (2000)
9. Marshall, S.P.: Identifying cognitive state from eye metrics. Aviation, Space, & Environmental Medicine **78**(5), 165–175 (2007)
10. Klingner, J., Kumar, R., Hanrahan, P.: Measuring the task-evoked pupillary response with a remote eye tracker. In: Proc. ETRA 2008 (2008)
11. Donoho, D.L.: Denoising by soft-thresholding. IEEE Transactions on Information Theory **41**(3), 613–627 (1995)
12. Welch, P.D.: The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms. IEEE Trans. Audio Electroacoustics **AU–15**, 70–73 (1967)