

An Evaluation of Pupillary Light Response Models for 2D Screens and Virtual Reality HMDs

Brendan John*
University of Florida
brendanjohn@ufl.edu

Arunava Banerjee
University of Florida
arunava@cise.ufl.edu

Pallavi Raiturkar*
University of Florida
pallaviraiturkar@ufl.edu

Eakta Jain
University of Florida
ejain@cise.ufl.edu

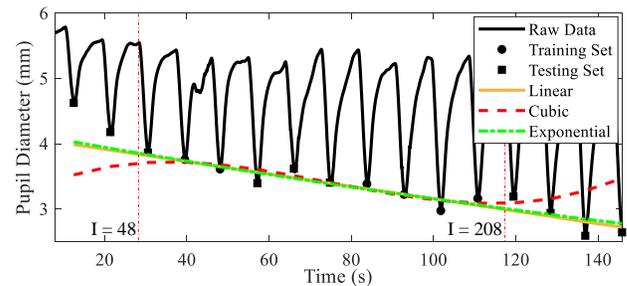
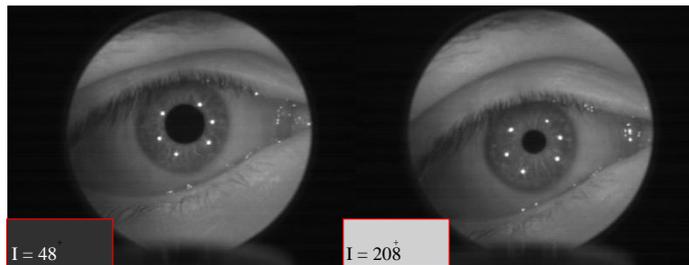


Figure 1: Left: Grayscale intensities $I = 48$ and $I = 208$ (insets) with corresponding eye image showing pupil size. Right: Linear, cubic, and exponential models fit to eight training points from VR pupillary light response calibration of one participant.

ABSTRACT

Pupil diameter changes have been shown to be indicative of user engagement and cognitive load for various tasks and environments. However, it is still not the preferred physiological measure for applied settings. This reluctance to leverage the pupil as an index of user engagement stems from the problem that in scenarios where scene brightness cannot be controlled, the pupil light response confounds the cognitive-emotional response. What if we could predict the light response of an individual's pupil, thus creating the opportunity to factor it out of the measurement? In this work, we evaluate three models of pupillary light response in 2D, and in a virtual reality (VR) environment. Our results show that either a linear or an exponential model can be fit to an individual participant with an easy-to-use calibration procedure. This work opens several new research directions in VR including modeling light response in 360° videos, training and simulation.

*Both authors contributed equally.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

VRST '18, November 28-December 1, 2018, Tokyo, Japan

© 2018 Association for Computing Machinery.

ACM ISBN XXXXXXXXXXXXXXXX...\$XX.00

<https://doi.org/XX.XXX/XXXXXXXX.XXXXXXX>

CCS CONCEPTS

• **Human-centered computing** → *User studies*; • **Applied computing** → *Psychology*;

KEYWORDS

Eyetracking, pupil dilation, light response, videos, virtual reality

ACM Reference Format:

Brendan John, Pallavi Raiturkar, Arunava Banerjee, and Eakta Jain. 2018. An Evaluation of Pupillary Light Response Models for, 2D Screens and Virtual Reality HMDs. In *VRST 2018: 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18)*, November 28-December 1, 2018, Tokyo, Japan. ACM, New York, NY, USA, 10 pages. <https://doi.org/XX.XXX/XXXXXXXX.XXXXXXX>

1 INTRODUCTION

The pupil is an opening in the eye that allows light to reach the retina. This opening changes size on an almost continuous basis because of factors ranging from pupillary unrest [52], brightness changes [53], sounds [43, 54], to emotional changes [8, 10] and cognitive workload changes [4, 5, 33, 40]. Changes in pupil size, typically measured as pupil diameter changes, are useful indicators of changing user engagement or mental workload, in human-computer interaction tasks such as personalized education or personalized narratives. While the pupil has been long studied [5, 9, 10], the effect of virtual reality (VR) environments on pupil diameter has yet to be explored.

The growth of virtual reality hardware and software in recent years has led to an increase in developers, artists, and researchers creating and exploring immersive virtual environments. Research

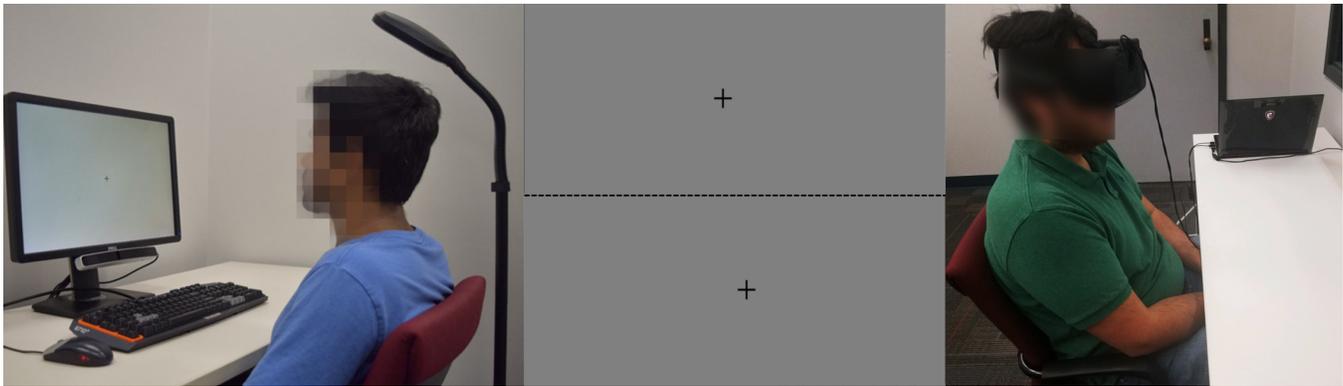


Figure 2: Left: 2D experimental setup. Center: Left eye view (Top) and right eye view (Bottom) of grayscale image for calibration. The black crosshair spans 2.5° visual angle at 4 meters distance. Right: VR experimental setup.

has shown that these environments create engaging experiences that transport users to exotic locations [24], provide different perspectives on social issues [44], and create training simulations in ways that 2D content simply cannot compare [3]. These applications can utilize pupil diameter as an alternative to self reports and other bio-physiological sensors for reliably measuring user engagement and cognitive load.

An algorithm that modifies an educational task or a game level based on the pupil diameter of the user needs to know if the pupil constricted because the scene was bright or because the user was no longer as engaged as before. The dilation or constriction of the pupil in response to brightness changes is known as Pupillary Light Response (PLR). Raiturkar et al. [46] demonstrated that a simple linear model could be used to fit PLR after a short 2D calibration procedure. In this paper we evaluate three pupillary light response models (linear, cubic, and exponential) with the goal of finding the best model for 2D and VR. Predictive error of each model determines which model performs best.

Modeling PLR has implications for both 2D and VR content as removing the influence of light on pupil diameter within task environments leaves a signal indicating cognitive load, emotional response, stress, and at a higher level, engagement. Several applications use bio-physiological sensors for such measurements through heart rate variability (HRV), galvanic skin response (GSR), monitoring brain signals with electroencephalogram (EEG), and through muscle responses measured by electromyography (EMG). GSR and EEG are particularly noisy signals that are sensitive to body movements, hard to capture in natural environments, and can require post-hoc analysis and processing. Pupil diameter is a non-invasive sensor and is becoming more accessible through eye trackers that are integrated within VR displays [1, 2]. Producing models that decouple pupil diameter from external factors, like light and depth, is critical to making this signal a viable index of engagement for this domain.

Our contributions can be summarized as follows:

- (1) Experimental design: In contrast to previous work that tests pupillary response models on static stimuli, we have created an ecologically valid test paradigm (Section 4.1).

- (2) Calibration procedure: We have proposed and tested a calibration framework for VR environments (Section 5.1). This framework builds upon previous literature accounts for VR specific factors such as depth of the calibration surface.

- (3) Model evaluation: We have evaluated three models of pupillary light response in both 2D and VR. Our experiments found the cubic model to be unsuitable, and the linear and exponential models to be similar in terms of error and ease of use (Sections 4 and 5).

2 RELATED WORK

When modeling the human eye as a camera, the pupil serves as an aperture stop controlling the amount of light hitting the retina. Pupil diameter is controlled by a reflex in the iris where a dilator muscle pulls outwards on the iris to make the opening bigger in dark environments, while circular muscles contract and shrink the pupil in bright lighting to reduce the amount of light entering [35]. Intrinsically photosensitive retinal ganglion cells respond to light, controlling the pupil diameter in an effort to ease adaptation to sudden changes in illumination [39, 58]. This is defined as Pupillary Light Reflex. The speed of pupillary constrictions from PLR are faster than that of dilations, and scale with stimulus intensity [19, 22]. PLR is the main function of varying pupil diameter, however other factors such as age, elasticity of the iris, wavelength of stimulus light, and iris color [6, 35, 57, 60] also influence pupil diameter. Specifically, external (auditory, visual) [8, 43, 54] and internal (emotional, cognitive, intentional) [5, 10, 40, 56] inputs are known to modulate pupillary size. Apart from these sensory reflexes, the pupil also shows regular constriction and dilation known as hippus or pupillary unrest [52].

Though pupillometry has been an area of research for a number of decades, the availability of commodity eye-trackers in the last decade has made it particularly accessible to researchers in different domains. Pupil diameter changes have been used as an index of engagement to score ups and downs in the course of a narrative video [30]. In application scenarios such as computer usage and driving, researchers have reported pupil diameter to be preferred as an index of user engagement than galvanic skin response [48] or heart rate variability [21] because pupil size measurement did not

require a physical connection to the participant, and they detected engagement changes with fewer participants compared to heart-rate variability. However, these works did not consider the impact of light intensity changes.

In order to test hypotheses around the effect of different factors, previous works have constructed tasks such that only one factor is at play at a time, for example, by matching luminance across stimuli images in pleasant versus unpleasant pictures [10], or, by using math problems as a 'neutral' workload task [40]. As a result, there is relatively sparse literature on how the various factors interact [50]. For real-world tasks, the interaction between factors is unavoidable. For example, while driving, the aggravation of being cut off will interact with the workload of changing lanes as well as with brightness changes because the sun went behind a cloud. Though pupil diameter changes have been reported to be informative of workload changes in the context of driving [21], other researchers have also reported that brightness related changes could obscure mental workload related changes [34]. Similarly, light intensity changes are unavoidable while watching movies or while viewing real-world scenes. A method to factor out light response from pupil diameter measurements would facilitate the use of pupil diameter as an index of emotional engagement, stress and cognitive load in a real-world setting. While there are several other factors to consider before applying pupil diameter as an index of workload or affect [12], we first focus on decoupling light reflex from pupillary measurements.

Over the last few decades, several pupillary light reflex models have been proposed. Most of these models predict the average pupil size as a function of incident light intensity, relying on experimental measurements [18, 41, 47, 59], while acknowledging that individuals vary and require personalized models. More recently, Pamplona et al. [42] proposed a physiologically-based model for pupil light reflex, described by a delay-differential equation. Watson and Yellott [53] provided a concise review of seven published PLR models, and also proposed a unified formula incorporating the effects of luminance, size of the adapting field, and age of the observer. All of these studies have been performed in a traditional 2D setting rather than in Virtual Reality environments.

VR displays have grown in accessibility, and use binocular vision to produce convincing 3D visuals using a high resolution near eye display. Several components of binocular viewing also influence pupil diameter [23, 35]. The components that make up stereo viewing include binocular disparity derived from the difference in left and right eye images [16], accommodation of the crystalline lens within the eye [15], the angle for which our eyes converge to focus on a particular depth [13], and visual fatigue [51]. Recently, Bharadwaj et al. [7] have shown that playing a video at decreasing depths will cause greater constrictions in pupil diameter, as objects move from far to near. Our pupil constricts in response to close objects to modify the point spread function of light, resulting in sharper retinal images [14, 55]. While pupil constrictions are a direct result of focusing on nearby objects and near vision [37], the exact contributions of accommodation and convergence have not been effectively modeled. One well established issue in virtual and augmented reality displays is termed vergence/accommodation conflict, created when our eyes converge on a virtual point at a distance, while our lens must be accommodated for the near eye display that is at a much closer depth [26, 29].

PLR modeling has only been applied in one instance to a 3D VR environment [17]. PLR was modeled using a cubic polynomial fit function. PLR was calibrated using a light meter installed inside the HMD, and a virtual room was increased in brightness from dark to bright (white), using nine total intensities. Model response was subtracted from measured pupil diameter to score emotional response to different emotional VR scenes. Change in pupil diameter from the PLR baseline was converted to percentage change, and compared between the different emotional stimuli. The largest dilation was found in the disgust emotion, with the happy emotion showing the least change from the baseline. While this work applies a PLR model in a VR environment, stimuli depth was not controlled as part of the experiment.

Alternate physiological sensors suffer from practical problems. Heart-rate is impacted by muscle movements, and this noise is increased in both EKG pads and finger clamps in case of imperfect placement [21]. Eye movements introduce artifacts into EEG [38], and GSR is sensitive to motion artifacts, temperature changes, as well as subject to strong attenuation over the course of minutes [31]. This tendency to habituate limits the usefulness of GSR for quantifying emotional engagement of a viewer over the time of prolonged media exposure, especially when the goal is to compare responses to different sections (e.g. early or late segments in a movie or VR experience). Pupil diameter responses co-vary strongly with GSR, but are not subject to habituation over time, and are insensitive to temperature and motion artifacts [9]. Past research has also shown that multimodal features are more reliable than a single modality [49], likely because each modality has its unique sources of noise. However, it is burdensome for a user to wear a multitude of sensors. If we limit ourselves to only a single modality, then pupil diameter has the clear advantage that it can be collected 'for free' along with gaze data. With a robust method to account for brightness changes we will be able to apply pupil diameter as a measure of emotive-cognitive state of the user in 2D and VR environments.

3 PUPILLARY LIGHT RESPONSE MODELS

Three models of PLR were evaluated. These linear, cubic, and exponential models are simple to fit with a least squares optimization and provide a simple personalized model for PLR. In both 2D and VR pupil diameter is recorded in response to changes between a black screen and grayscale image, as part of a calibration procedure.

Linear Model. In [46], a linear model of pupillary light response was presented: for an intensity I , the pupil diameter d is a constant d_1 times I , with an offset δ . More precisely,

$$d(t) = d_0 + d_1 \cdot I(t - \delta) \quad (1)$$

The intensity value $I(t)$ is in the range $[0, 255]$. In [46], the delay parameter δ , is tuned to each participant. This model does not account for non-linearity (the pupil cannot dilate indefinitely). Yet, the simplicity of the model makes it very easy to calibrate parameters to each individual participant by minimizing the error between the prediction of the linear model for input grayscale intensity values, and the recorded pupil diameter.

Cubic Model. A cubic model for PLR can be represented as,

$$d(t) = d_0 + d_1 \cdot I(t - \delta) + d_2 \cdot I^2(t - \delta) + d_3 \cdot I^3(t - \delta). \quad (2)$$

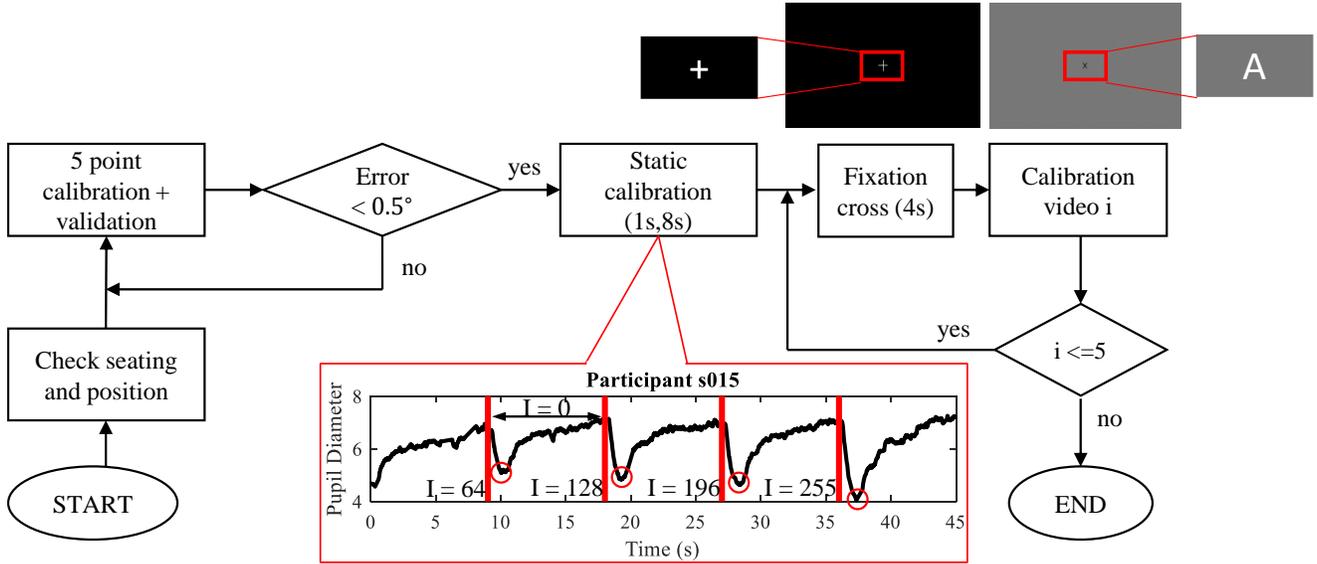


Figure 3: 2D Experimental procedure. Participants went through a five-point calibration, a light response calibration, and then watched the five test videos. Inset shows pupillary response of one participant to the light response calibration.

This model can account for non-linearity within the training data, but has potential for overfitting [53]. A cubic fit has a tendency to diverge outside the range of training intensities.

Exponential Model. Holladay [27] proposed an exponential formula to predict pupil diameter based on luminance,

$$d(t) = a \cdot e^{b \cdot I(t-\delta)}. \quad (3)$$

This model can capture non-linearity at low and high intensities, with its monotonic nature reducing error outside the range of training data. However, at a small scale the fitted exponential curve can revert to a mostly linear fit.

4 EVALUATION ON 2D SCREEN

We recorded gaze positions and pupil diameters while participants went through a light response calibration, and subsequently viewed five grayscale intensity videos. Two videos were short (60-90 seconds) and three videos were long (approximately five minutes each). We describe the experiment protocol, summarize the PLR models evaluated, and present an evaluation of the trained models on the grayscale intensity videos.

4.1 Protocol

Participants went through an informed consent procedure in accordance with an IRB approved protocol. Each participant was seated in a comfortable chair whose height was adjusted to that participant. A white lamp was placed behind the participant, and all other lights were switched off. The data collected consisted of two sessions with a break in between.

Participants went through a five-point gaze calibration procedure, followed by a five-point validation. Gaze calibration was

repeated till the average validation error was below 0.5° visual angle. Then, each participant was presented with the first brightness change calibration: a set of slides of increasing grayscale intensities ($I = 0, 64, 128, 192, 255$). Each calibration slide was displayed for one second, followed by a black slide ($I = 0$) for eight seconds (see Figure 3 (inset)). Following the break, we ran the gaze calibration again. We also ran the brightness calibration with the grayscale intensity slides again. Next, the participant was presented with the grayscale intensity test videos (V_1-5). These videos had a resolution of 1680×1050 and contained uniform slides of varying grayscale intensities in the range 0-255. As part of another experiment in which participants viewed narrative videos, we generated grayscale videos matching the grayscale intensity in a 2° foveal neighborhood of the participant's gaze position (see Supplementary Document). As a result, each participant saw a different set of five videos. The videos had durations between 63s and 335s.

Because the test videos contained no semantic content, it was quite boring for a participant, and we found that one of our pilot participants fell asleep. We added a randomly chosen letter (A-Z) to the center of each frame in the calibration video, sized at 20×20 pixels, i.e., less than 1° , and black in color. The letter was changed every second. The participant was tasked with saying the letter out loud. We also added a soft humming bird tone to the filler slide to focus the participant before the start of a trial. In total, there were five test videos for each participant. We randomized the order of test videos, and each video was preceded by a black slide with a white fixation cross, displayed for 4s. The procedure followed (post break) is presented in Figure 3.

Participants viewed the stimuli on a 22 inch (1680×1050) monitor. They sat approximately 60-65 cm away from the screen. One degree visual angle was around 37 pixels at this distance. All data was recorded on a SensoMotoric Instruments (SMI) RED-m eyetracker

with SMI's Experiment Center software at 120 Hz. Raw data including timestamps, gaze position (x,y) and pupil diameter (mm) were exported as a comma separated file using SMI's BeGaze software. A MATLAB script assigned each data item to the video frame with closest time stamp. Figure 2 (Left) shows our experimental setup.

Twenty four participants (17 female, age range 18 to 24 years) were recruited from a participant pool available through the university, in accordance with an IRB approved protocol. Participants were students who received course credit in return for their participation. Participant ethnicity was 67% White/Caucasian, 13% Black/African-American, 8% Hispanic/Latino, 8% Asian/Pacific-Islander, and 4% Other. All participants had normal or corrected-to-normal vision. Our eye-tracker could calibrate participants with contact lenses, but not eye-glasses. Female participants were instructed not to wear eye makeup as mascara interferes with pupil tracking.

4.2 Data Processing

We discarded trials if more than 50% samples were missing or corrupted. For example, one participant fell asleep and was removed from subsequent data analysis. One participant did not finish data collection because the eye-tracker lost connection to the computer. For some participants, a subset of videos had more than 50% of samples missing, so we removed those videos from analysis. Finally the number of participants available was 19, 18, 15, 20, and 20 for V_1, V_2, V_3, V_4, V_5 respectively. Next, all data points where the gaze position was outside the stimulus presentation screen were discarded. Samples with velocity greater than the peak pupil constriction velocity, i.e., 5.65mm/s were also removed [11]. We then linearly interpolated any samples that were missing (pupil diameter was zero). Finally, we smoothed the signal using a low-pass 5th order Butterworth filter with cut off frequency 10Hz [32]. All data processing was done in MATLAB R2017a.

Because participants underwent the light response calibration twice, we picked the session in which the pupillary constrictions were less noisy. If both sessions had clean data, we picked the first calibration session. For each participant, we manually marked out the pupil minima following each of the four training intensities ($I = 64, 128, 192, 255$). We then fit the linear, cubic and exponential models for each participant by a linear least squares optimization using MATLAB functions `pinv()`, `fit()` and `polyfit()` respectively.

Using these models, we predicted the pupillary response for each participant on every test video. The predicted pupil diameter and the measured pupillary diameter was centered on the mean for comparison. We selected the ten sharpest spikes and ten biggest dips based on video grayscale intensity as sampling points, representing pupil constrictions and dilations respectively. Finally, we computed the Root Mean Squared Error (RMSE) between the predicted pupillary response of each model and the measured pupillary response at these sample points. Lower RMSE values indicate a better prediction. We also computed RMSE values by predicting pupil diameter on the entire test video.

4.3 Results

Figure 4 (Left) shows the measured pupil diameter for two participants on the video V_1 and the corresponding grayscale intensity. For both participants, there is a dip in the pupil diameter in the first

Video	Linear		Cubic		Exponential	
	μ	σ	μ	σ	μ	σ
V_1	0.45	0.11	0.52	0.11	0.47	0.12
V_2	0.46	0.10	0.51	0.11	0.47	0.11
V_3	0.55	0.10	0.69	0.14	0.56	0.10
V_4	0.62	0.12	0.72	0.15	0.63	0.12
V_5	0.59	0.15	0.72	0.20	0.59	0.15
Overall	0.53	0.14	0.63	0.18	0.55	0.14

Table 1: Means(μ) and Standard Deviations(σ) of RMSE in mm for five test video predictions, across all participants for three models of pupillary light reflex.

few seconds after stimulus onset. This dip is because participants are shown a black fixation cross slide prior to the start of the trial. When the stimulus video starts to play, the participants' pupils constrict. The black dotted curve marks the grayscale intensity for each frame. Because the videos were generated on a per-participant basis, their pupils were subject to different brightness values. In each case, the pupil constricted for bright frames (e.g., frame #950) and dilated for dark frames (e.g., frame #1050).

We computed the Pearson correlation coefficient between the video grayscale intensity and measured pupil diameter for each participant and each test video. Because the pupil constricts in response to brighter frames, the correlation coefficients will be negative. Figure 4 (Right) shows the box plots of these correlation coefficients. As a baseline, we computed the correlation coefficient between the grayscale intensity observed by a participant and the pupil diameter of a different randomly chosen participant, for the same video. The correlation coefficient for the same participant ($M = -0.53, SD = 0.26$) is closer to -1 than for a different participant ($M = -0.34, SD = 0.28$). The group means are significantly different (one tailed independent samples t-test, $t(91) = -5.02, p < 0.05$).

Figure 5 (Right) illustrates the measured pupil diameter for test video V_1 , from one of our participants. Overall RMSE values for the five videos are reported in Table 1. The linear model has the least error of 0.53mm for predicting pupil diameter with the exponential coming in a close second at 0.55mm. Errors were not significantly different between the two models. Table 2 contains the means and standard deviations of RMSE for the ten pupil constrictions and dilations of each model across participants and videos. Across all test videos, the linear model has the least error of 0.42mm for constrictions and 0.6mm for dilations. RMSE values are consistently lower for constrictions compared to those of dilations.

4.4 Discussion

We performed a light response calibration consisting of 4 grayscale intensity slides, shown to participants in flashes of 1 second, with 8 seconds of a black slide in between. The purpose of the calibration procedure was to train individual models to predict changes in pupil diameter in response to a continuous grayscale intensity video. Figure 4 (Right) shows that pupil responses were significantly different between participants for the same intensities.

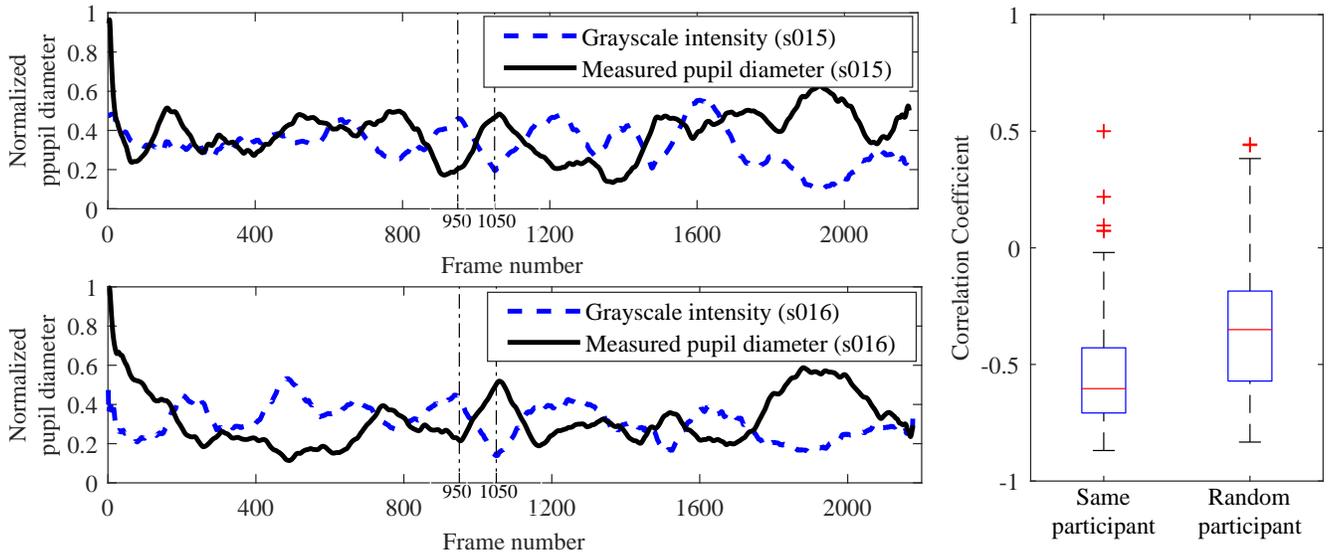


Figure 4: Left: Measured pupil diameter (solid black) on test video V_1 for two participants (s015,s016). The average grayscale intensity (dotted blue line) is different for the two participants because the videos were generated on a per-participant basis. Right: Boxplot of correlation coefficients between the grayscale intensity observed by a participant and the pupil diameter of the same participant, and a different randomly chosen participant.

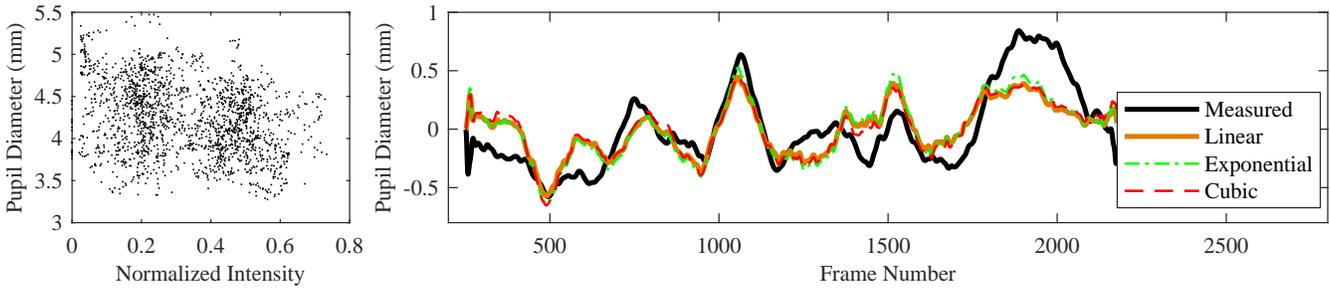


Figure 5: PLR model testing. Left: Pupil diameter as a function of normalized grayscale intensity for the video V_1 . Right: Pupillary response of a participant (s016) to the test video V_1 (black), and corresponding model predictions of pupil diameter. Pupil diameter data was centered on the mean.

The linear model had the lowest RMSE for predicting changes in pupil diameter during the test videos. RMSE values for constrictions were on average $0.2mm$ less than predicted dilations. The RMSE values were within one standard deviation of the measured pupil diameter. These results are promising, as a generalized model built strictly from grayscale intensities does not capture the relationship between pupil diameter and brightness (Figure 5 Left). A short calibration that elicits pupillary constrictions is able to predict pupillary response to brightness changes.

We observed that all three models predict that the pupil diameter is generally larger at a given grayscale intensity, compared to the observed pupil diameter. However, we noticed that the relative changes predicted by the model matched the pupil diameter closely over the course of the video. To account for this we subtracted

the mean from both predicted and measured pupil diameter before computing RMSE, as seen in Figure 5 (Right). Based on our evaluation a simple linear model is sufficient to predict PLR while viewing dynamic stimuli on 2D screens by predicting constrictions and dilations. The next section evaluates if these PLR models can be applied in a VR environment.

5 EVALUATION ON VR HEAD MOUNTED DISPLAYS

In VR the screen is located much closer to the eye and covers a larger field of view than in 2D, meaning the amount of light entering the eye from the stimulus is greater. Figure 6 illustrates that illuminance, or luminous flux per area, is around three or four times greater as measured in VR compared to 2D.

Video	Linear		Cubic		Exponential	
Constriction	μ	σ	μ	σ	μ	σ
V_1	0.35	0.11	0.46	0.16	0.37	0.12
V_2	0.34	0.10	0.50	0.29	0.35	0.11
V_3	0.46	0.17	0.59	0.22	0.48	0.19
V_4	0.47	0.24	0.54	0.23	0.49	0.25
V_5	0.46	0.28	0.57	0.27	0.49	0.30
Overall	0.42	0.18	0.53	0.23	0.44	0.19
Dilation	μ	σ	μ	σ	μ	σ
V_1	0.45	0.14	0.54	0.17	0.49	0.16
V_2	0.54	0.16	0.63	0.20	0.59	0.19
V_3	0.57	0.15	0.68	0.22	0.60	0.17
V_4	0.79	0.23	0.82	0.25	0.81	0.24
V_5	0.65	0.23	0.76	0.33	0.70	0.26
Overall	0.60	0.18	0.69	0.23	0.64	0.20

Table 2: Means (μ) and Standard Deviations (σ) of RMSE in mm across all participants. Predictive error was computed using the 10 most prominent constrictions and dilations in the five test grayscale intensity videos.

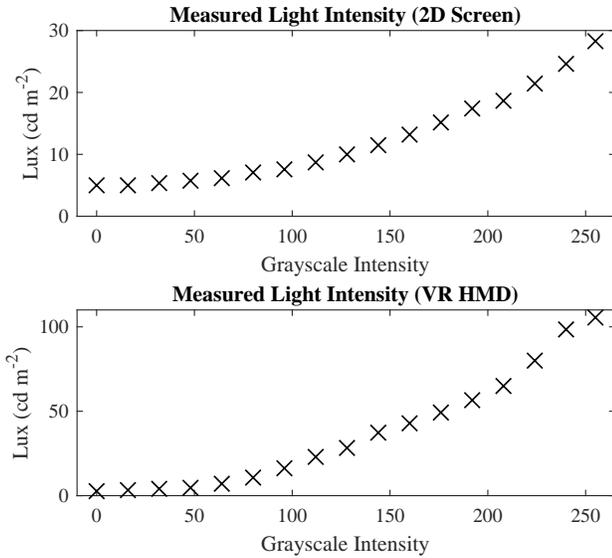


Figure 6: Measured illuminance of each intensity image from 2D calibration procedure (Top), and from the VR calibration procedure (Bottom). Light was measured using a Sekano light meter in the approximate eye position 60 cm away from the screen in the 2D setup, and 23 mm away from the left eye lens of the VR HMD. Lux values are larger in the VR condition for the same grayscale value.

The calibration procedure from Section 4.1 was translated to VR using 16 different grayscale intensities at one depth. Depth was inferred from stereo convergence of the eyes and binocular disparity between the left and right eye images. Please see the Supplementary for more details. Five participants were shown intensities in

ascending order from dark to bright, while five participants viewed intensities in a randomized order. This modification in presentation sequence was to determine if a non-monotonic order establishes a more generalized PLR model, due to temporal effects or adaptation to bright stimuli. We evaluate the predictive error of each model across training sets of different sizes, to determine which works best in a VR environment, and to establish how many calibration points are necessary for an optimal fit.

5.1 Protocol

All participants went through an informed consent procedure in accordance with an IRB approved protocol. Figure 2 (Right) shows how each participant was seated in a swivel chair for the experiment. An Oculus-DK2 HMD with built-in 60Hz binocular SMI eye tracker was used to record gaze position and pupil diameter of each eye. Unlike in Section 4, there was no lamp needed in the room, as most external lighting is blocked out when wearing the HMD. The lights in the room were switched off prior to the light response calibration.

Sixteen different intensities were shown for one second each ($I = 16, 32, 48, \dots, 255$), followed by eight seconds of a black image ($I = 0$). Participants saw one of two calibration sequences, either with the intensities presented in ascending order from 16 to 255, or randomly shuffled. The grayscale image is set to the specified intensity with a black crosshair in the center, spanning 2.5° visual angle, as seen in Figure 2 (Center). The image was placed 4 meters directly in front of the viewer, with gray values filling the rest of the screen. The HMD was programmed not to respond to head movements, so the crosshair remained centered in the field of view at all times.

Data was collected from ten participants (2 female, age range from 18 to 30) recruited from a university environment. Participants were mostly students, who received course credit in return for their participation. Participant ethnicity was 60% White/Caucasian, 10% Black/African-American, 20% Hispanic, and 10% Native American. All participants had normal or corrected-to-normal vision. Our eye tracker could accurately calibrate participants with contact lenses, but not eye-glasses.

5.2 Model Training & Evaluation

Each participant logged $M = 16$ pupil constrictions, extracted similarly to Section 4.2. Training sets consisting of four, six, eight, and ten input intensities were used for each participant for least squares fitting using MATLAB functions *pinv()*, *fit()* and *polyfit()* respectively. The remaining points were used as the testing set. 50 randomly generated sets were used for each size, with RMSE computed between the model predicted and measured pupil diameter at the testing intensities.

5.3 Results

Table 3 and Figure 7 contain the means and standard deviations of RMSE for each model and training sets in the ascending and shuffled groups. Averaged across all training sets, the exponential model has the lowest error of $0.27mm$. Figure 8 illustrates model fits for each size testing set. The cubic model has an error over $1mm$ for $N = 4$ as a cubic function overfits the data with fewer data points.

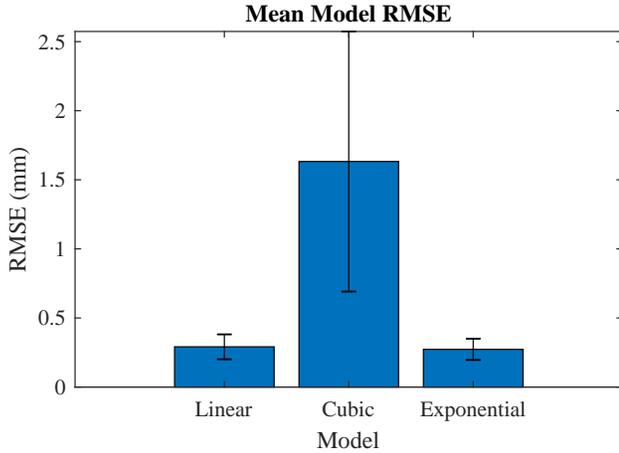


Figure 7: Means and standard deviations of predictive RMSE in mm for three PLR models averaged across all training set sizes. Training sets of size four, six, eight, and ten were used.

Training Set	Linear		Cubic		Exponential	
	μ	σ	μ	σ	μ	σ
Ascending						
$N = 4$	0.42	0.10	4.47	3.30	0.36	0.10
$N = 6$	0.31	0.07	0.48	0.22	0.29	0.08
$N = 8$	0.33	0.07	0.32	0.18	0.30	0.07
$N = 10$	0.27	0.08	0.21	0.10	0.25	0.08
Shuffled						
$N = 4$	0.31	0.12	6.39	3.48	0.30	0.09
$N = 6$	0.25	0.09	0.44	0.18	0.24	0.07
$N = 8$	0.25	0.09	0.34	0.10	0.24	0.07
$N = 10$	0.23	0.07	0.23	0.07	0.22	0.05
Overall	0.29	0.09	1.63	0.94	0.27	0.08

Table 3: Means (μ) and Standard Deviations (σ) of testing RMSE in mm across all participants for three models of PLR applied to the VR calibration. Training sets of four different sizes were used as described in Section 5.2. Predictive error was computed using intensities not in the training set.

5.4 Discussion

We evaluated models that predict pupil diameter as a function of grayscale intensity in VR at a fixed depth. We measured the predictive error of three models (linear, cubic, exponential), and found the exponential model performed best. The calibration procedure from 2D was modified by increasing the number of sampled intensities, and introducing a randomized ordering sequence of presented intensities. The grayscale intensities and corresponding fixation cross was presented at a depth of 4 meters in the virtual environment.

Across all three models we found that our average predictive RMSE fell within the range of $0.22mm$ and $6.4mm$. Considering only $N \geq 8$ the maximum error is $0.34mm$ falling within the range of previously reported 2D results [46]. The largest error was created due to cubic model overfitting the data when only a few points were used for training. With test sets of size $N \geq 8$ the average

RMSE of the three models fell within one standard deviation of recorded pupil constrictions for each participant. Only one participant had predictive error that was higher than this standard deviation. Averaged across each participant the standard deviation of pupil diameter was $0.47mm$, meaning errors for the linear and exponential model were within this range. Overall, the exponential model errors were not significantly different from the linear model. Figure 7 shows that the exponential curve ultimately looks linear, with the model allowing some curvature. This suggests that PLR can be modeled using either linear or exponential fits, but that the cubic model is prone to significant errors with less than eight training points.

The PLR calibration can be performed by presenting the grayscale intensities in ascending order, or in a randomized sequence. Table 3 shows that the exponential model generalized best across all training sets ($\mu = 0.27mm$). Performance of models when we used a shuffled or ascending calibration sequence were comparable, suggesting that either method could be used.

These results confirm that we can model pupillary response using existing models from 2D viewing. Our method is a first step towards applying PLR models in VR, and motivates the need to evaluate them in more practical scenarios like 360° videos, or 3D scenes. Our calibration stimuli were presented at one depth, with binocular disparity between the left and right eye view (Figure 2, Center).

The presence of other depth cues such as depth of focus [25], shadows [45], shading, and perspective may have an effect on pupil diameter. We know that pupil diameter correlates with convergence and accommodation, with the pupil constricting as stimuli moves closer to the viewer [7]. Future work could incorporate stimuli at different depths to build a more complex model of pupil diameter. Additionally, pupil diameter recorded while viewing continuous grayscale videos in VR would allow for a more rigorous evaluation of the model, with respect to temporal effects.

At a single depth our proposed method can be used to predict pupillary light response in VR. Applications such as measuring emotional response during 360° video viewing [36], monitoring cognitive load in virtual teaching environments [28], or measuring Quality of Experience (QoE) in VR [20] can apply this method to decouple PLR from pupillary measurements. Currently systems use bio-physiological sensors that measure HRV and GSR, and would benefit from switching to pupil diameter to reduce the usage of noisy and invasive sensors that interfere with user experience and comfort. VR and AR is the future of interactive computing, user experience, and entertainment, of which the pupil provides a window into the user's true affective and cognitive state once confounding factors are modeled and removed.

5.5 Limitations

The models evaluated in this paper do not consider factors such as accommodation, the speed at which the pupillary system can change diameter, and the temporal dynamics of the response. We are currently using an instantaneous PLR model, i.e. the model does not incorporate past values of pupil diameter. Our model has another shortcoming: While viewing a relatively bright portion of video may cause a pupillary constriction, the same brightness value could later cause the pupil to be slightly larger due to adaption,

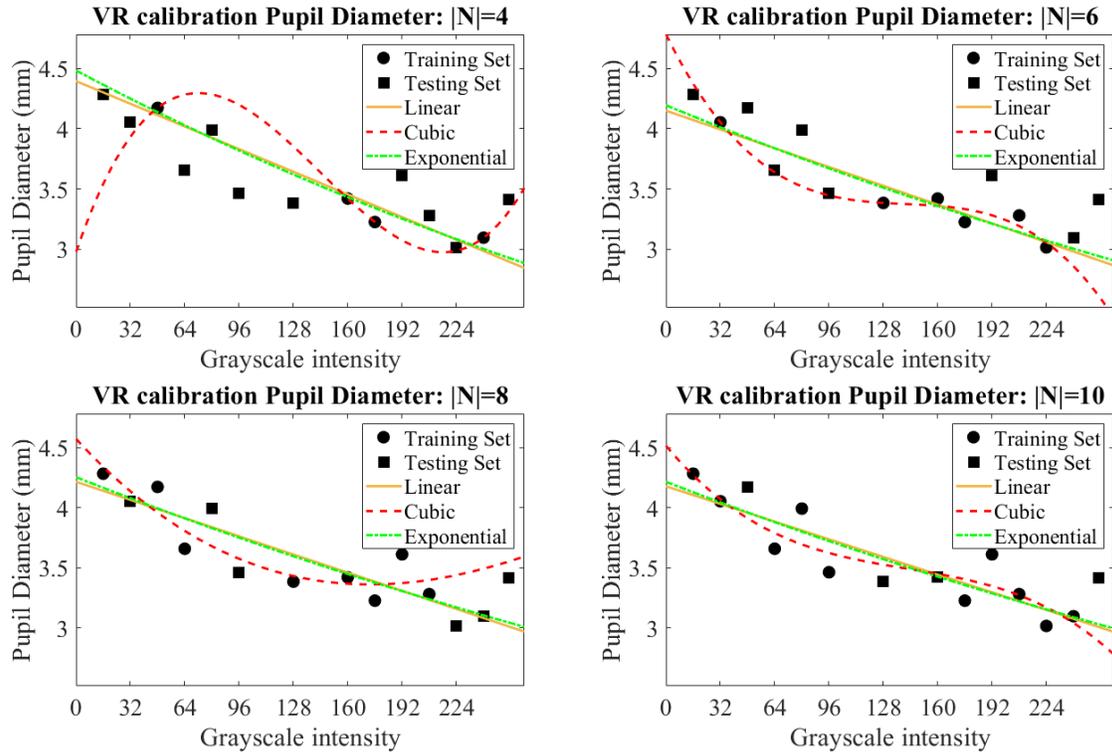


Figure 8: Visualization of the three PLR models tested with data from one participant (s003) in VR with a randomized presentation order of intensities. Each plot corresponds to a random training set of size N used to generate a least squares fit.

which would not be captured by our model. While these factors are no doubt important, learning parameters from data is subject to the curse of dimensionality. Adding more factors would require increasingly longer calibration procedures. The more burdensome a calibration procedure is, the less likely a user is to adopt it. Our proposed paradigm seeks to make the calibration as quick and easy as possible.

Another limitation of this work is that data was collected on displays within a particular range of illuminance (Figure 6). The peak brightness of newer VR displays, such as the HTC Vive, can exceed 200 Lux. Depending on the display used, the most accurate model of PLR may change due to the logarithmic response of our pupil to brightness. Regardless, the presented calibration procedure provides a method to identify the optimal model to predict pupil diameter in response to grayscale intensity values.

6 CONCLUSION

We have presented a method consisting of an individual calibration and constriction-based models of pupil diameter in 2D and VR. Our 2D results suggest that using a linear model we can predict constrictions within $0.41mm$ and dilations within $0.60mm$ when viewing grayscale videos. In VR we have found that an exponential model predicts pupil diameter with the lowest error ($0.27mm$), replicating findings in 2D. We believe this method can be applied to current

applications within VR to remove pupillary light response. With a reliable method of fitting PLR models in VR, future work can focus on decoupling light response from pupillary measurements to ultimately render pupil diameter a faithful measure of affective and cognitive state.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Award No. IIS-1566481. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- [1] [n. d.]. Apple acquires SMI. <http://fortune.com/2017/06/27/apple-acquires-eye-tracking-tech-firm-for-augmented-reality/>. ([n. d.]). Accessed: 2018-08-15.
- [2] [n. d.]. Google acquires Eyefluence. <https://www.businessinsider.com/google-buys-eyefluence-vr-ar-eye-tracking-startup-2016-10>. ([n. d.]). Accessed: 2018-08-15.
- [3] Gunnar Ahlberg, Lars Enochsson, Anthony G Gallagher, Leif Hedman, Christian Hogman, David A McClusky, Stig Ramel, C Daniel Smith, and Dag Arvidsson. 2007. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *The American journal of surgery* 193, 6 (2007), 797–804.
- [4] Brian P Bailey and Shamsi T Iqbal. 2008. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction (TOCHI)* 14, 4 (2008), 21.

- [5] Jackson Beatty. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin* 91, 2 (1982), 276.
- [6] Oliver Bergamin, Andreas Schoetzau, Keiko Sugimoto, and Mario Zulauf. 1998. The influence of iris color on the pupillary light reflex. *Graefes archive for clinical and experimental ophthalmology* 236, 8 (1998), 567–570.
- [7] Shrikant R Bharadwaj, Jingyun Wang, and T Rowan Candy. 2011. Pupil responses to near visual demand during human visual development. *Journal of vision* 11, 4 (2011), 6–6.
- [8] M. M. Bradley. 2009. Natural selective attention: orienting and emotion. *Psychophysiology* 46 (2009), 1–11.
- [9] Margaret M Bradley and Peter J Lang. 2015. Memory, emotion, and pupil diameter: Repetition of natural scenes. *Psychophysiology* 52, 9 (2015), 1186–1193.
- [10] 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.
- [11] Fion D Bremner. 2012. Pupillometric Evaluation of the Dynamics of the Pupillary Response to a Brief Light Stimulus in Healthy Subjects. *Investigative ophthalmology & visual science* 53, 11 (2012), 7343–7347.
- [12] Anne-Marie Brouwer, Thorsten O Zander, Jan BF van Erp, Johannes E Korteling, and Adelbert W Bronkhorst. 2015. Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Frontiers in neuroscience* 9 (2015), 136.
- [13] JF Cardenas-Garcia, HG Yao, and S Zheng. 1995. 3D reconstruction of objects using stereo imaging. *Optics and Lasers in Engineering* 22, 3 (1995), 193–213.
- [14] WN Charman and Helen Whitefoot. 1977. Pupil diameter and the depth-of-field of the human eye as measured by laser speckle. *Optica Acta: International Journal of Optics* 24, 12 (1977), 1211–1216.
- [15] W Neil Charman and Hema Radhakrishnan. 2009. Accommodation, pupil diameter and myopia. *Ophthalmic and Physiological Optics* 29, 1 (2009), 72–79.
- [16] Chien-Chung Chen and Christopher William Tyler. 2015. Shading beats binocular disparity in depth from luminance gradients: Evidence against a maximum likelihood principle for cue combination. *PLoS one* 10, 8 (2015), e0132658.
- [17] Hao Chen, Arindam Dey, Mark Billinghurst, and Robert W. Lindeman. 2017. Exploring Pupil Dilation in Emotional Virtual Reality Environments. In *ICAT-EGVE 2017 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, Robert W. Lindeman, Gerd Bruder, and Daisuke Iwai (Eds.). The Eurographics Association. <https://doi.org/10.2312/egve.20171355>
- [18] SG De Groot and JW Gebhard. 1952. Pupil size as determined by adapting luminance. *JOSA* 42, 7 (1952), 492–495.
- [19] CJ Ellis. 1981. The pupillary light reflex in normal subjects. *British Journal of Ophthalmology* 65, 11 (1981), 754–759.
- [20] Martin Ervik. 2017. *Contributing Factors of Immersion in Interactive Digital Storytelling-Evaluation of Heart Rate as a Parameter for Immersion in QoE assessment of Interactive Storytelling*. Master's thesis. NTNU.
- [21] Thomas M Gable, Andrew L Kun, Bruce N Walker, and Riley J Winton. 2015. Comparing heart rate and pupil size as objective measures of workload in the driving context: initial look. In *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 20–25.
- [22] HS Gradle and Walter Ackerman. 1932. The Reaction Time of The Normal Pupil: Second Communication. *Journal of the American Medical Association* 99, 16 (1932), 1334–1336.
- [23] Daniel G Green, Maureen K Powers, and Martin S Banks. 1980. Depth of focus, eye size and visual acuity. (1980).
- [24] Daniel A Guttentag. 2010. Virtual reality: Applications and implications for tourism. *Tourism Management* 31, 5 (2010), 637–651.
- [25] Robert T Held, Emily A Cooper, and Martin S Banks. 2012. Blur and disparity are complementary cues to depth. *Current biology* 22, 5 (2012), 426–431.
- [26] David M Hoffman, Ahna R Girshick, Kurt Akeley, and Martin S Banks. 2008. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision* 8, 3 (2008), 33–33.
- [27] LL Holladay. 1926. The fundamentals of glare and visibility. *JOSA* 12, 4 (1926), 271–319.
- [28] Hsiu-Mei Huang, Ulrich Rauch, and Shu-Sheng Liaw. 2010. Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach. *Computers & Education* 55, 3 (2010), 1171–1182.
- [29] Anke Huckauf. 2005. Virtual and real visual depth. In *Proceedings of the 2nd symposium on Applied perception in graphics and visualization*. ACM, 172–172.
- [30] Harish Katti, Karthik Yadati, Mohan Kankanhalli, and Chua Tat-Seng. 2011. Affective video summarization and story board generation using pupillary dilation and eye gaze. In *Multimedia (ISM), 2011 IEEE International Symposium on*. IEEE, 319–326.
- [31] JL Kenemans, MN Verbaten, W Sjouw, and JL Slangen. 1988. Effects of task relevance on habituation of visual single-trial ERPs and the skin conductance orienting response. *International journal of psychophysiology* 6, 1 (1988), 51–63.
- [32] Jeff Klingner, Rakshit Kumar, and Pat Hanrahan. 2008. Measuring the task-evoked pupillary response with a remote eye tracker. In *Proceedings of the 2008 symposium on Eye tracking research & applications*. ACM, 69–72.
- [33] Arthur F Kramer. 1991. Physiological metrics of mental workload: A review of recent progress. *Multiple-task performance* (1991), 279–328.
- [34] Andrew L Kun, Oskar Palinko, and Ivan Razumenic. 2012. Exploring the effects of size and luminance of visual targets on the pupillary light reflex. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 183–186.
- [35] Yves Le Grand. 1957. *Light, colour, and vision*. Wiley.
- [36] Benjamin J Li, Jeremy N Bailenson, Adam Pines, Walter J Greenleaf, and Leanne M Williams. 2017. A Public Database of Immersive VR Videos with Corresponding Ratings of Arousal, Valence, and Correlations between Head Movements and Self Report Measures. *Frontiers in psychology* 8 (2017), 2116.
- [37] Irene E Loewenfeld and Otto Lowenstein. 1993. *The pupil: Anatomy, physiology, and clinical applications*. Vol. 2. Wiley-Blackwell.
- [38] Juan-Miguel López-Gil, Jordi Virgili-Gomá, Rosa Gil, and Roberto García. 2016. Method for Improving EEG Based Emotion Recognition by Combining It with Synchronized Biometric and Eye Tracking Technologies in a Non-invasive and Low Cost Way. *Frontiers in computational neuroscience* 10 (2016).
- [39] Emma L Markwell, Beatrix Feigl, and Andrew J Zele. 2010. Intrinsically photosensitive melanopsin retinal ganglion cell contributions to the pupillary light reflex and circadian rhythm. *Clinical and Experimental Optometry* 93, 3 (2010), 137–149.
- [40] Sandra P Marshall. 2002. The index of cognitive activity: Measuring cognitive workload. In *Human factors and power plants, 2002. proceedings of the 2002 IEEE 7th conference on*. IEEE, 7–7.
- [41] Parry Moon and Domina Eberle Spencer. 1944. On the stiles-crawford effect. *JOSA* 34, 6 (1944), 319–329.
- [42] Vitor F Pamplona, Manuel M Oliveira, and Gladimir VG Baranoski. 2009. Photorealistic models for pupil light reflex and iridal pattern deformation. *ACM Transactions on Graphics (TOG)* 28, 4 (2009), 106.
- [43] 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.
- [44] Keshav Prasad, Kayla Briët, Obiageli Odimegwu, Olivia Connolly, Diego Gonzalez, and Andrew S Gordon. 2017. "The Long Walk" From Linear Film to Interactive Narrative. In *AIIDE 2017 Workshop on Intelligent Narrative Technologies*, Vol. 10.
- [45] Antonio Medina Puerta. 1989. The power of shadows: shadow stereopsis. *JOSA A* 6, 2 (1989), 309–311.
- [46] Pallavi Raiturkar, Andrea Kleinsmith, Andreas Keil, Arunava Banerjee, and Eakta Jain. 2016. Decoupling light reflex from pupillary dilation to measure emotional arousal in videos. In *Proceedings of the ACM Symposium on Applied Perception*. ACM, 89–96.
- [47] Prentice Reeves. 1918. Rate of pupillary dilation and contraction. *Psychological Review* 25, 4 (1918), 330.
- [48] P Ren, A Barreto, Y Gao, and M Adjouadi. 2011. Comparison of the use of pupil diameter and galvanic skin response signals for affective assessment of computer users. *Biomedical sciences instrumentation* 48 (2011), 345–350.
- [49] Mohammad Soleymani, Maja Pantic, and Thierry Pun. 2012. Multimodal emotion recognition in response to videos. *IEEE transactions on affective computing* 3, 2 (2012), 211–223.
- [50] Robert F Stanners, Michelle Coulter, Allen W Sweet, and Philip Murphy. 1979. The pupillary response as an indicator of arousal and cognition. *Motivation and Emotion* 3, 4 (1979), 319–340.
- [51] Kazuhiko Ukai and Peter A Howarth. 2008. Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations. *Displays* 29, 2 (2008), 106–116.
- [52] Shiro Usui and Lawrence Stark. 1982. A model for nonlinear stochastic behavior of the pupil. *Biological Cybernetics* 45, 1 (1982), 13–21.
- [53] Andrew B Watson and John I Yellott. 2012. A unified formula for light-adapted pupil size. *Journal of Vision* 12, 10 (2012), 12–12.
- [54] Michael W Weiss, Sandra E Trehub, E Glenn Schellenberg, and Peter Habashi. 2016. Pupils dilate for vocal or familiar music. *Journal of Experimental Psychology: Human Perception and Performance* 42, 8 (2016), 1061.
- [55] Gerald Westheimer. 1964. Pupil size and visual resolution. *Vision research* 4, 1-2 (1964), 39–45.
- [56] Olivier White and Robert M French. 2017. Pupil diameter may reflect motor control and learning. *Journal of motor behavior* 49, 2 (2017), 141–149.
- [57] Barry Winn, David Whitaker, David B Elliott, and Nicholas J Phillips. 1994. Factors affecting light-adapted pupil size in normal human subjects. *Investigative ophthalmology & visual science* 35, 3 (1994), 1132–1137.
- [58] JM Woodhouse and FW Campbell. 1975. The role of the pupil light reflex in aiding adaptation to the dark. *Vision research* 15, 6 (1975), 649–653.
- [59] Francis A Young and William R Biersdorf. 1954. Pupillary contraction and dilation in light and darkness. *Journal of comparative and physiological psychology* 47, 3 (1954), 264.
- [60] Rockefeller SL Young and Eiji Kimura. 2008. Pupillary correlates of light-evoked melanopsin activity in humans. *Vision research* 48, 7 (2008), 862–871.