Is the Avatar Scared? Pupil as a Perceptual Cue

Anonymous Authors

Abstract

The importance of eyes for virtual characters stems from the intrinsic social cues in a person's eyes. While previous work on computer generated eyes has considered realism and naturalness, there has been little investigation into how details in the eye animation impact the perception of an avatar's internal emotional state. We present three large scale experiments (N≈500) that investigate the extent to which viewers can identify if an avatar is scared. We find that participants can identify a scared avatar with 75% accuracy using cues in the eyes including pupil size variation, gaze, and blinks. Because eye trackers return pupil diameter in addition to gaze, our experiments inform practitioners that animating the pupil correctly will add expressiveness to a virtual avatar with negligible additional cost. These findings also have implications for creating expressive eyes in intelligent conversational agents and social robots.

Keywords: perception, virtual avatars, animated characters, eye tracking, pupil, gaze

1 Introduction

Eyes are a critical piece for truly compelling virtual avatars. "Oculus animi index", in other words, the eye is the window to the soul. Artists carefully create detailed appearance models for their characters' eyes, and spend hours on keyframing to have them move just right [1]. There is much work on algorithms and techniques to create eyes that feel alive [2, 3, 4]. Research has advanced the appearance of eyes to the point of photo-realistic [5, 6, 7] and matching advances have been made on the animation of the eyes [8].

Eyes are typically animated as eyeballs that are rigged to follow a desired target, or to match eye-tracking data from an actor. Previous research has focused on making the dynamics of the eyeball realistic [9], as well as the motions of the eyelids [10, 11]. With advances in eye tracking, there is now work on adding subtle pupil dynamics, such as pupil jitter [12, 13] and pupil light reflex [14]. While this past research has performed perceptual studies to show that adding realistic dynamics to the eyeballs, eyelids, and pupils improves the perceived realism and naturalness of a virtual avatar, there has been little investigation into how cues in the eyes impact the perceived internal state of a virtual avatar. We therefore aim to answer the question: do cues as subtle as pupil size impact how the avatar is perceived. Such an investigation is relevant because each cue adds potential complexity to the modeling of a virtual avatar.

Pupil dilation is an indicator of stress and emotional arousal [15, 16] as well as cognitive load [17, 18]. As a result, the pupil has been used in computational analysis, for example as a physiological index for affective video processing [19, 20]. While the variation in pupil size is a useful feature for computational processing, our motivation for this work is to determine whether it is an intrinsic social cue that *humans* use to make judgments about the avatar. We designed three studies to answer our research questions. First, can naive viewers identify if an avatar is scared from only the information contained in its eyes? Second, can naive viewers make this judgment even if gaze direction and blink rate are not available as a cue? Third, is correct pupil dynamics a critical perceptual cue? In this paper, our contributions are as follows:

1. We are the first study to investigate in depth how subtle social cues in the eyes communicate the internal emotional state of an avatar, as different from prior art that investigates the impact of eye detail on an avatar's realism, naturalness, personality, and trustworthiness. We focus our inves-

tigation on an emotion that can be reliably elicited in a lab setting: being scared while watching a movie involving, for example, a zombie attack.

2. We report a sequence of large scale studies (N≈500) that investigate the role of subtle cues in an avatar's eyes for communicating if this avatar is scared.

3. We find that a cue as subtle as pupil size variation is perceivable to humans, in addition to gaze and blinks. These findings underscore the importance of high fidelity pupillary dynamics for expressive avatars in animated movies, conversational agents, and social virtual reality.

2 Related Work

Increasing realism and naturalness has been a long term goal when it comes to virtual characters. Much research has gone into generating realistic and natural movements for the eyes as presented in Ruhland et al.'s survey [8]. Generating gaze, i.e., the trajectory of the eyeball, has, for example, been studied from the point of view of replaying a user's gaze position [21], as well as automatically animating the eyes as well as hand and body movements when catching an object [9]. Researchers have also proposed methods to animate the subtle movement of the eyelids while rolling and folding [11] and blinking [10].

There has also been research into creating realistic models of the interior parts of the eye. Pamplona et al. [14] modeled iris deformation as the pupils dilate and constrict. Duchowski et al. [22] modeled eyes motions by adding noise to gaze and pupil diameter. Bérard et al. [6] demonstrated a capture system that could accurately reconstruct areas such as the sclera, cornea, and iris individually, and combine the parts into a complete eye model. They extended the work further to a parametric model for eye reconstruction from a single photo [7].

This body of work on the fine details of the eyes is driven in part by the belief that unrealistic eye motions are responsible for the uncanniness of computer generated faces [23]. In the same vein, studies have shown that lack of facial expression in the upper parts of the face (including eye brows and eyes) would increase the "uncanniness" of a character for certain emotions [24]. Researchers have also investigated our perception of eye motions particularly and which types of information we gather from an avatar's eyes. MacQuarrie and Steed studied the perception of gaze direction of volumetrically captured avatars [25]. They found that gaze based viewing direction judgements became less accurate when the avatar was looking at targets further away from the user. Steptoe et al. [26] have shown that high fidelity visual representation of avatars will increase accuracy with which people identify the avatar's direction of gaze, together with subjective reports of authenticity.

Beyond the gaze direction, an avatar's eyes also impact the viewers' judgments of the characters personality and emotions. Examples are the character's truthfulness [27] and perceived trustworthiness and aggressiveness [28]. Pupil size is particularly interesting as a social cue, as, in contrast to eye motions, it cannot be faked or controlled [29]. Pupil diameters are affected by arousing image stimuli, audio stimuli and VR scenes, regardless of the valence of the stimuli [15, 30, 31, 32]. While changes in pupil size appear to be a very subtle cue, researchers have found that viewers can notice them. Viewers judged partners with larger pupils to be more positive and attractive compared to partners with smaller pupils [33, 34]. Heterosexual individuals also showed larger pupils when looking at pictures of the opposite sex [35]. Researchers found that one person's pupil size would mimic the pupil size of another person during close social interaction [36] or when shown photos of another person [37]. Harrison et al. [37] also found out that small pupils were rated more intense for sad emotion. However, it did not affect other emotions such as happy, angry or neutral. In this work, we investigate how the avatar's pupil dynamics - specifically, time-averaged pupil size and its variation - impact viewers' judgement of the avatar being scared.

Impact As high fidelity avatars are critical to social virtual reality, there is active research on faithfully reproducing facial motion, eyes, and lips for the particular case of avatars driven by a human user wearing a virtual reality headset [38, 39, 5]. Eye trackers are a necessary sen-

sor for social virtual reality as that is the only way to reproduce the user's gaze and create eye contact. While eye trackers are already being built into VR head-mounted displays [40, 41], and pupil diameter is easily computed from the same video stream, should we go through the effort of recreating the correct pupillary movements for an avatar's eyes? Or is it too subtle to be perceivable? Our work addresses this open question. Our findings inform practitioners that pupillary dynamics should not be ignored and that the expressiveness of an avatar's or virtual character's eyes can be improved with the addition of detailed pupillary motion, which can also be applied to conversational agents and social robots.

3 Experimental Framework

Overview We designed a sequence of three studies with a forced choice task that became progressively harder. *Study 1* required participants to identify which avatar was watching a scary movie when one avatar's eyes were animated (gaze, pupil and blink) with data collected during a scary movie viewing session and the other avatar's eyes were animated with data collected during a grayscale slides viewing session. Each grayscale slide had the same brightness as the corresponding frame in the scary movie. *Study 2* required participants to make the same judgement, with the following differences in the study conditions: while the first avatar stays the same, the other avatar's eyes were animated with matching gaze and blink data from the scary movie viewing session as the first avatar, but pupil diameter from the grayscale slides viewing session. In *Study 3*, the other avatar's eyes were animated with gaze data from the movie viewing session but pupil diameter using procedurally generated jitter. This sequence of studies allowed us to probe whether identifying the scared avatar was possible at all (*Study 1*), when gaze and blinks were made identical (*Study 2*), and when the average size of the pupil was also made identical but the dynamics were still different (*Study 3*).

Eye-tracking Data We used the dataset of Raiturkar and colleagues to obtain eye-tracking data for people watching a scary movie using an SMI RED-m eye tracker with 120 Hz. Their dataset contains eye motions from ten people watching excerpts from movies and YouTube videos[1]. Details of the data collection protocol are available in the associated publication [19]. We selected an excerpt from a zombie movie (*Decay*) that is freely available online under a Creative Commons license. Screenshots from the movie can be found in Figure 1 in Raiturkar et al.'s work [19], The ten second excerpt we selected contained a zombie attack on the heroine[2]. This attack was marked by Raiturkar et al. as an arousal eliciting event in their publication. We agreed with their assessment of this event being the scariest zombie attack. From the set of available eye motions, we selected those in which the pupils changed in diameter by more than 1.5mm in the three seconds following the zombie attack (ID: *s004*, *s009* and *s015*). This is an empirical threshold that we chose because the change beyond 1.5mm is visually noticeable for viewers. The selected eye motions were from people who had self-reported an engagement score greater than or equal to 7, when asked to rate their experience on a scale of 1 to 10 (1=least engaging). Finally, we checked that the recorded pupil diameter had less than 40% missing samples due to capture devices or blinking. As this is the first study of its kind, we selected these participants to determine if respondents could perceive cues in the pupil when those cues are amply present. The raw pupil diameter recordings from the three eye motions are shown in Figure 1.

We obtained additional eye-tracking data from the same people as above. The difference was that this time they were watching a video comprising only grayscale slides without any semantic content or indicator of gaze direction. Thus we ensured that in this case there was no reason to expect that they were scared. The grayscale videos were generated as follows: for each person whose eye tracking data was collected, the foveal neighborhood of their gaze position in each frame of the zombie video was

[1] Dataset "Gaze & Pupil Diameter III" available at `http://jainlab.cise.ufl.edu/publications.html#GAZEPUPIL`

[2] The time stamps are 66279ms to 76575ms relative to the longer excerpt available within the dataset of Raiturkar et al. [19]

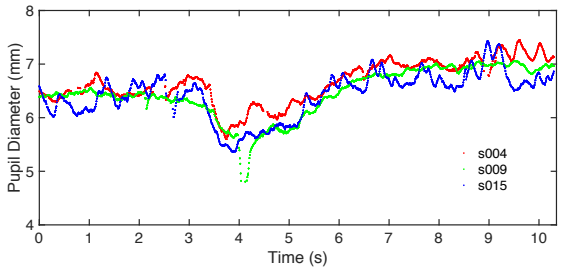

Figure 1: Raw pupil diameter recordings from the three selected eye motions of people watching the scary movie.

identified. The color values of the pixels in this neighborhood were converted into an average grayscale intensity. A corresponding frame was generated for the neutral video such that its grayscale value matched this average intensity. A similar protocol was published by John et al. [42].

Avatar Creation We adopted the highly realistic human avatar from Jörg et al. [12]. The avatar's eyeball movement was driven by a point in a 2-D plane through a point constraint to mimic someone watching a screen. We adjusted the plane's size and distance from the character to match the real world eye tracking setup. As our virtual character is not at the scale of a real character, we used its iris size compared to the average iris size of a real person as a reference to scale the pupil sizes to match the character.

The captured data (gaze coordinates and pupil size) was manually cleaned by an artist. We detected blinks in the eye-tracking data using the values of the pupil diameters. If the pupil diameter dropped to zero at certain frames, then we identify those frames as blinks. We removed the incorrect pupil data by deleting the affected keyframes so that the frames before and after the blinks would be interpolated. We used the same method to remove blinks in the gaze data. The scene was set up in Maya 2016 and rendered with Solid Angle's Arnold 5 renderer. The lighting was adjusted and the avatar was rendered in a close-up view focusing on the eye area (Figures 2, 3, and 4). This view displays the eye movement in high detail. We used a resolution of 854 $\times$ 480 with a frame rate of 24 frames/second. The stimuli for our three experiments can be seen in the provided video.

Perceptual Study Protocol We ran our data collection under a protocol approved by the uni-

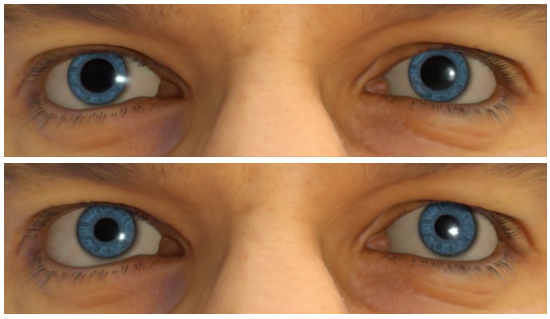

Figure 2: Rendering of avatar *s015* from one of our eye motions is shown here for the conditions Scary (Top) and Grayscale (Bottom) in Study 1.

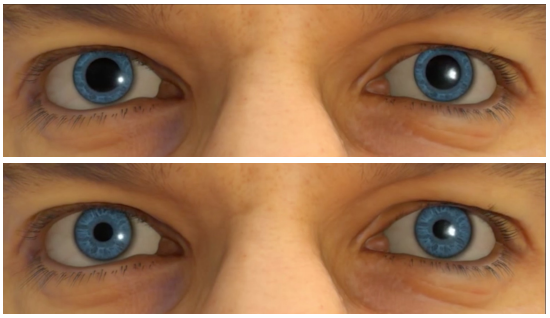

Figure 3: Rendering of avatar *s015* from one of our eye motions is shown here for the conditions Scary (Top) and Neutral (Bottom) in Study 2.

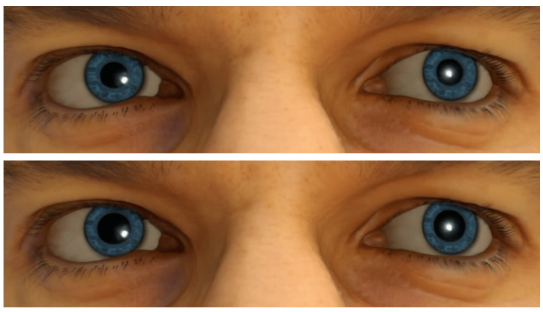

Figure 4: Rendering of avatar *s015* from one of our eye motions is shown here for the conditions Scary (Top) and Procedural (Bottom) in Study 3.

versity review board (IRB). Respondents were redirected to Survey Gizmo to take an online survey. First, we asked demographics questions including age, gender and ethnicity. Then survey respondents were shown the stimuli videos. They were asked to watch each video in its entirety. Because each video was only 10.3 seconds long, we looped it four times and created a stimulus that was 41.2 seconds long to give survey respondents enough time to observe the details. We implemented a timer that constrained the survey respondents to spend at least 35 seconds on each page. The stimuli were presented in randomized order. Every participant watched all six of the stimuli. We asked viewers to use a large screen such as a laptop or desktop to view this survey with the browser window maximized.

Respondents were asked two questions after each stimulus presentation. For the question "One of these people is watching a scary movie. Which one is it?", their response was recorded via a radio button with two options: "Top" or "Bottom". For the question "How difficult did you find this task?", the response was recorded on a 7-point Likert scale with 1(very easy) and 7(very difficult). We adopted the forced choice paradigm provided by signal detection theory as a solution to measure the detection rate of a stimulus independent of respondents' individual internal bias. The advantage of this paradigm is that the measured detection rates are independent of internal motivations such as social desirability or the amount of money respondents are being paid. Though users in actual application scenarios will likely not encounter digital avatars side-by-side, our study design establishes if a signal (in this case, pupil dynamics) is detectable.

At the end, we asked two supplemental questions. The first question was multiple-choice: "How often do you play video games?" Five options were provided: "0 hours/week", "0-7 hours/week", "7-14 hours/week", "14-21 hours/week", "Above 21 hours/week". The second question was "How often do you watch animated movies or TV shows?" The options were the same as for the first supplemental question. Finally, survey respondents were asked a free-form text question: "What cue(s) did you use to say whether the person is watching a scary movie?" All the questions in the survey were mandatory. The data privacy guidelines required by the authors' institutional Review Board were followed and all the surveys are anonymous. Participants' worker IDs were not associated with their survey responses.

4 Study 1: Identify The Scared Avatar

In Study 1, we investigated if an avatar with the captured eye motions of people watching scary movies would be perceivably different from the same avatar with eye motions from people watching only light intensity changes. Our motivation was to establish if respondents can identify the scared avatar.

4.1 Description

The independent variable in this study was the video used to elicit the eye motions: a scary zombie video (Condition Scary) versus a video composed of only grayscale slides that changed in intensity value (Condition Grayscale), see Figure 2. The gaze directions, blinks, and pupil diameters came from the corresponding eye-tracking data (Figure 5, 8).

Stimuli videos generated from the same person were shown synchronized and aligned vertically. To control for order effects, we created stimuli videos with Condition Scary on top and Condition Grayscale at the bottom, as well as Condition Grayscale on top and Condition Scary at the bottom. In total, each participant watched six different stimuli videos (3 eye motions $\times$ 2 orders).

4.2 Results

Demographics We recruited 104 survey respondents from Amazon Mechanical Turk. One survey respondent entered the wrong survey code. 4 survey respondents entered the code without taking the survey. We received 99 complete survey responses. Respondents were compensated 0.8 dollars when they completed the survey. We excluded those responses where the IP addresses were repeated to have a total of 88 unique respondents. Our survey respondents were well

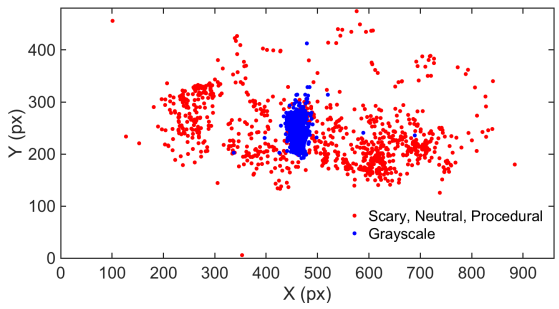

Figure 5: The numeric values of gaze positions from eye motion *s015* are graphed for the different experimental conditions in Study 1.

distributed by age (age range 19-72, μ = 31.89, σ =9.34), gender (32 female) and ethnicity (25 Asian, 49 White, 6 Black or African American, 2 Hispanic or Latino, 5 American Indian or Alaska Native, 1 Others).

Validation We created two additional stimuli videos from different eye motions of a person (*s010*). These were presented such that the top and the bottom videos are identical. Each pair was presented twice, resulting in four dummy stimuli sets. Respondents were asked to respond to two easy questions for each stimuli set, one question from the Top/Bottom pair, and one question from the glasses/eye color pair. The questions were: "The person on the bottom is watching a scary movie. Click "Bottom"."., "The person on top is watching a scary movie. Click "Top"."., "What's the eye color of this person?" (Blue/Black) and "Is this person wearing glasses?" (Yes/No). We only kept the data from a survey respondent if they got >= 6 questions correct. As a result, 76 respondents were kept for the remainder of the analysis.

Recognition Accuracy For the remaining 76 survey respondents, we computed the number of correct responses as follows: we coded each row with the actual answer; if it matched the respondent's answer, then the number of correct responses was incremented. We computed the percentage of correct responses for all six stimuli videos for each respondent. Figure 6 illustrates the distribution of correct responses as a histogram. Table 1 shows the percent correct responses broken down by Eye Motion ID.

We used the statistical analysis package R to further analyze the data. The distributions of percentage correct responses were not normal

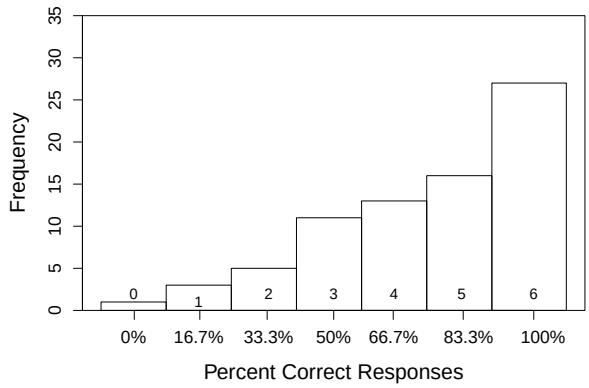

Figure 6: Study 1 results. Histogram of Percent Correct Responses for consistent survey respondents. The number of correct responses is shown inside the bars of the histogram.

(Shapiro-Wilk, W=0.86, $p < 0.001$), as seen in Figure 6. A one-sample Wilcoxon signed-rank test reported that the percent of correct responses (μ = 74.56%, σ = 25.88%) is significantly above chance (V = 1953.5, $p < 0.001$). We also performed a permutation test to account for the number of stimuli and number of participants [43]. Based on this permutation test1, the empirical chance level is 55.4% for a p-value = 0.05 and the probability of correct response of 50%. Hence, our result in Study 1 (74.56%) is significantly above chance. We computed effect sizes as Pearson's r correlation (z-score divided by square root of number of observations), r = 0.66. The effect size is a large effect (r = 0.5) [44]. These results demonstrate that survey respondents are able to distinguish the avatar that was watching a scary movie from the avatar that was watching a baseline neutral video.

Breakdown by Eye Motion ID We examined if respondents' accuracy in identifying the scared avatars was biased in favor of any individual person whose eye tracking data was used to generate the stimuli videos. We computed the mean and standard deviation of the percentage correctness by each Eye Motion ID in Table 1. There was no significant difference between the percentages of correct responses for each Eye Motion ID. A non-parametric Friedman test of repeated measures was conducted (Friedman χ^2 value of 1.11, $p = 0.57, p > 0.05$). As a re-

sult, we retained all three eye motions when we generated stimuli videos for subsequent experiments.

5 Study 2: Pupil as a Cue

The human pupil changes diameter as part of the autonomic nervous system response to feeling scared or stressed. The pupil diameter also changes in response to changing scene brightness. If the pupil was simply changing size in response to changes in brightness, would survey respondents still think the avatar was scared? In Study 2, we investigated this question by controlling the gaze and blink data while manipulating the source of pupil diameter changes.

5.1 Description

The independent variable in this study was the source of the pupil diameter values: pupil diameter recorded while people were watching a scary movie (Condition Scary) versus pupil diameter recorded while people were watching a video composed of only grayscale slides that changed in intensity value (Condition Neutral), see Figure 3. The gaze directions and blink rates were controlled. These values came from the recordings where the people were watching a scary movie for both conditions. Note that Condition Scary in this study is identical to Condition Scary in Study 1.

Our goal was not to test which cue (gaze direction, blink, pupil) dominates, rather, the goal was to test whether having the *correct* pupil diameter makes a difference to how often the avatar is judged as being scared. By "correct", we mean pupil sizes and variations which are a result of an emotional response rather than light regulation. If the correct pupil diameter values make a difference to respondents' judgments, then the message to content creators would be: for truly expressive eyes, we recommend that they use the correct pupil diameters.

Avatars were rendered as before. We showed survey respondents the Condition Scary video and Condition Neutral video in a vertical layout. The videos were synchronized to play simultaneously. To account for order effects, each respondent also saw each video pair again, this time with the top and bottom reversed. The questions asked were the same as in the previous study. The study was conducted on Amazon Mechanical Turk under an IRB approved protocol. Our dependent variable was the response to "One of these people is watching a scary movie. Which one is it?" We calculated the metric Percent Correct Responses as before.

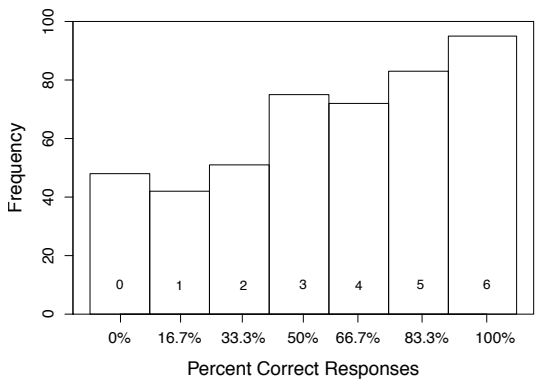

Figure 7: Study 2 results. Histogram of percent correct responses. The number of correct responses is marked inside each bar.

5.2 Results

Demographics In total, we paid 500 survey respondents \$0.8 for their time. We received 510 completed responses in Qualtrics. After eliminating survey respondents with repeated IP addresses, there were 500 distinct respondents left (223 female, age range 18-80 years, μ = 36.04 years, σ = 11.89 years, 97 Asian, 308 White, 38 Hispanic or Latino, 36 Black or African American, 9 American Indian or Alaska Native, 12 Others).

Validation We used the same validation procedure as in Study 1. 466 survey respondents were kept for the remaining analysis.

Recognition Accuracy For the remaining 466 survey respondents, we calculated the metric Percent Correct Responses.

The distribution of Percent Correct Responses by survey respondent was not normal (Shapiro-Wilk, W = 0.91, $p < 0.001$). The histogram in Figure 7 shows the distribution. As shown in the last row of Table 1, on average Condition Scary was selected as 'the person watching the scary movie' in 58.7% of the presen-

ID	Percent Correct Responses					
	Study 1		Study 2		Study 3	
	μ	σ	μ	σ	μ	σ
s004	76.31%	32.12%	45.28%	40.68%	58.88%	34.14%
s009	73.68%	35.09%	67.70%	39.53%	62.13%	35.34%
s015	73.68%	35.09%	63.20%	40.73%	58.31%	34.69%
Overall	74.56%	25.88%	58.72%	32.63%	59.77%	24.93%

Table 1: Means and standard deviation of percent correct responses broken down by Eye Motion ID for Study 1, 2, 3.

ID	Condition							
	Scary				Neutral			
	μ	σ	Max	Min	μ	σ	Max	Min
s004	6.66	0.40	7.45	5.6	5.41	0.34	6.12	3.75
s009	6.44	0.43	7.05	4.81	4.23	0.40	5.12	3.83
s015	6.47	0.44	7.43	5.46	4.43	0.28	5.13	3.83

Table 2: Study 2 results. Mean and standard deviation of pupil diameter values for each Eye Motion ID.

tations ($\sigma = 32.6\%$), whereas Condition Neutral was picked in 41.28% of the presentations ($\sigma = 32.6\%$). A one-tailed Wilcoxon signed-rank test found that Percent Correct Responses for Condition Scary (μ = 58.7%, σ = 32.6%), averaged across the 466 respondents is significantly above chance (V = 50209, $p < 0.001$).

The permutation test shows that the empirical chance level is 51.53% for p-value = 0.05 and the percent correct response being the random chance level (50%) [43]. Our result 58.7% is significantly above the empirical chance level.

We computed effect sizes as Pearson's r correlation r = 0.27. The effect size is close to a medium effect size (r = 0.3) [44]. Thus, we concluded that even if an avatar's gaze and blinks correspond to eye-tracking data while watching a scary movie, it matters where the pupil diameter values are drawn from.

Breakdown by Eye Motion ID We checked if respondents' judgments of whether Condition Scary was the 'person watching a scary movie' depended on the person whose data was used to generate that avatar. For each respondent and each Eye Motion ID, we aggregated the number of times Condition Scary was selected as the response. Table 1 shows the mean and standard deviation of percent correct responses for each condition broken down by Eye Motion ID. We found a main effect of Eye Motion ID (Friedman χ^2 = 135.03, $p < 0.001$). The mean values indicated that for *s004* Condition Neutral was selected as the "person watching a scary movie" slightly more often than Condition Scary. This could be because for some people (such as *s004*), gaze is the dominant cue rather than pupil. If we refer to the average pupil diameters in Table 2, we see that the pupil was on average much larger in Condition Scary relative to Condition Neutral for both *s009* and *s0015*. In contrast, the average pupil diameter for *s004* for the two conditions was more similar. In Study 3, we compared condition Scary to a baseline where the average pupil diameter was the same as the average pupil diameter in Condition Scary. We found that respondents could identify better than chance when s004 was scared. Details follow in the next section.

6 Study 3: Pupil Diameter Variation as a Cue

In Study 3, we investigated whether a pupil with the same average pupil diameter but a procedurally generated jitter would be as effective in conveying that the avatar is scared as replaying the originally recorded pupil diameter. In other

words, if we control for the average size of the pupil, can respondents pick up on the pupil diameter variation as a cue.

6.1 Description

The independent variable in this study was the source of the variation in the pupil diameter values: pupil diameter recorded while people were watching a scary movie (Condition Scary) versus pupil diameter generated procedurally (Condition Procedural), see Figure 4. The average pupil diameter was kept the same in both conditions (Table 3). The gaze directions and blink rates in both conditions were from eye tracking data when people were watching the scary movie.

Pink noise has been widely used to model biological motions [45]. Duchowski et al. [45] and Krejtz et al. [13] found that it can be used to model the subtle changes in pupil size. A similar result was found by Jörg et al. [46] showing that adding pink noise to the pupil size and eyeball motions of different virtual characters created animations with similar perceived naturalness than a motion captured approach and perceived as significantly more natural than animations with only keyframed eyeball motions and pupil size changes without pink noise. To create the procedural pupil, we generated pink noise by following the approach by Duchowski et al. [45]. We used MATLAB's Colored Noise function with an alpha value of 1 to generate the noise. We added an offset to the noise to make the mean pupil diameter the same as Condition Scary. We scaled the noise to result in an average amplitude of 0.0023, a value used in Jörg et al. [12]. They generated procedural pupil jitter for two motions: a reading motion and a motion inspired by an eye tracking calibration procedure where viewers looked nine points on a white board without any changes in light. We used the average noise amplitude for the nine point motion as that type of motion is most similar to our stimuli. The resulting pupil diameter values are illustrated in Figure 8. We showed survey respondents two videos at a time in vertical layout played simultaneously. To minimize order effects, we also created a corresponding video for each stimulus with the same avatars but reversed vertical layout.

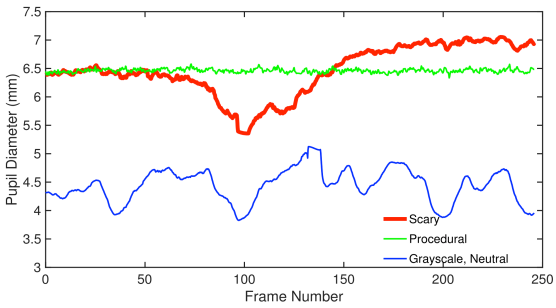

Figure 8: Pupil diameter values of Eye Motion ID *s015* used in Study 1, 2 and 3.

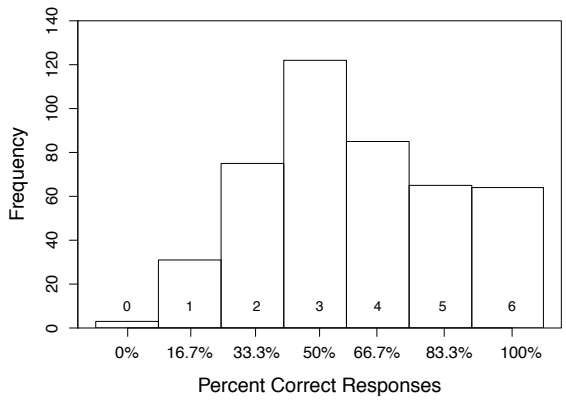

Figure 9: Study 3 results. Histogram of percent correct responses. The numbers inside the bars are the number of correct responses.

6.2 Results

Demographics Study 3 was conducted separate from Study 2. In total, we paid 515 survey respondents $0.8 to complete the study. After eliminating survey respondents with repeated IP addresses, there were 484 distinct respondents left (203 female, age range 18-70 years, μ = 35.11 years, σ = 11.09 years, 128 Asian, 282 White, 27 Hispanic or Latino, 30 Black or African American, 12 American Indian or Alaska Native, 5 Others).

Validation We used the same validation procedure as in Study 1 and Study 2. 445 respondents were kept for the remainder of the analysis.

Recognition Accuracy For each of the 445 respondents, we computed the metric Percent Correct Responses. If respondents identified the scared avatar as the one whose pupils were a replay of the original recording then this response

ID	Condition							
	Scary				Procedural			
	μ	σ	Max	Min	μ	σ	Max	Min
s004	6.66	0.40	7.45	5.98	6.65	0.03	6.76	6.54
s009	6.43	0.43	7.05	4.81	6.42	0.03	6.53	6.32
s015	6.47	0.44	7.43	5.46	6.46	0.03	6.58	6.33

Table 3: Study 3 stimuli generation: Mean and standard deviation of pupil diameter for each Eye Motion ID.

was marked as a correct response. If they selected the avatar whose pupils were generated procedurally, it was marked incorrect. Figure 9 shows the distribution of correct responses as a histogram. The empirical chance level given the number of stimuli and number of participants is 51.61% when the p-value = 0.05% and the percent correctness is 50% [43]. Our Percent Correct Responses (μ = 59.77%, σ = 24.93%) is significantly above the empirical chance level. Again, we computed effect sizes as Pearson's r correlation r = 0.27, which is close to a medium effect size (r = 0.3) [44].

Breakdown by Eye Motion ID Similar to Study 1 and Study 2, we examined respondents' accuracy in identifying the scared avatars. The breakdown of Percent Correct Responses by Eye Motion ID is shown in Table 1. We did not find a significant main effect of Eye Motion ID (Friedman χ^2 = 3.61, $p = 0.16 > 0.05$).

7 Respondents' Self-Reports

After each stimulus presentation, we asked respondents to self-report how difficult they found the task. At the end of the survey, we additionally asked respondents how often they play video games and how often they watch animated movies and shows. Finally, respondents were asked to write out in free-form text the cues they thought they used to answer whether the avatar was watching a scary movie. In this section, we analyze the self-reports for each study.

7.1 Task Difficulty

We examined survey respondents' self reports on the difficulty of the task (1=very easy to 7=very difficult). Figure 10 (a) shows that, generally speaking, survey respondents found the tasks neither too difficult nor too easy. The mean values trend upward as we move from Study 1 ($\mu = 3.81, \sigma = 1.51$) to Study 2 ($\mu = 4.19, \sigma = 1.49$) and then to Study 3 ($\mu$ = 4.79, σ = 1.60). This is consistent with the design of the experiments. In the first study, respondents have a number of cues to help them identify which avatar is scared, including gaze patterns, pupil diameter changes, and blink rate. In the second study, respondents only have the pupil as a cue because differences in gaze patterns and blink rate are controlled for, in effect making the task harder. In the third study, besides controlling gaze and blinks as in Study 2, we additionally controlled for the average size of the pupil, thus leaving only the change in pupil size as a cue, which makes the task even harder.

We ran individual Friedman rank sum tests for each study. In Study 1 and Study 3, we found no significant effect of Eye Motion ID (Study 1: Friedman χ^2 = 1.11, $p > 0.05$, Study 3: Friedman χ^2 = 3.61, $p > 0.05$). In Study 2, there was a significant effect of Eye Motion ID (Friedman χ^2 = 135.03, p $<$ 0.0001). Survey respondents reported that they found it more difficult to identify *s004* (μ = 4.49, σ = 1.55) than *s009* (μ = 4.04, σ = 1.69) and *s015* (μ = 4.03, σ = 1.68). This is consistent with the trend in Percent Correct Responses, and is explained by Table 2 – the average pupil diameters for the two conditions for *s004* were closer to each other than for the other two eye motions. This motivated the design for Study 3, where the average pupil diameters were made equal for both conditions (see Table 3 for numerical values).

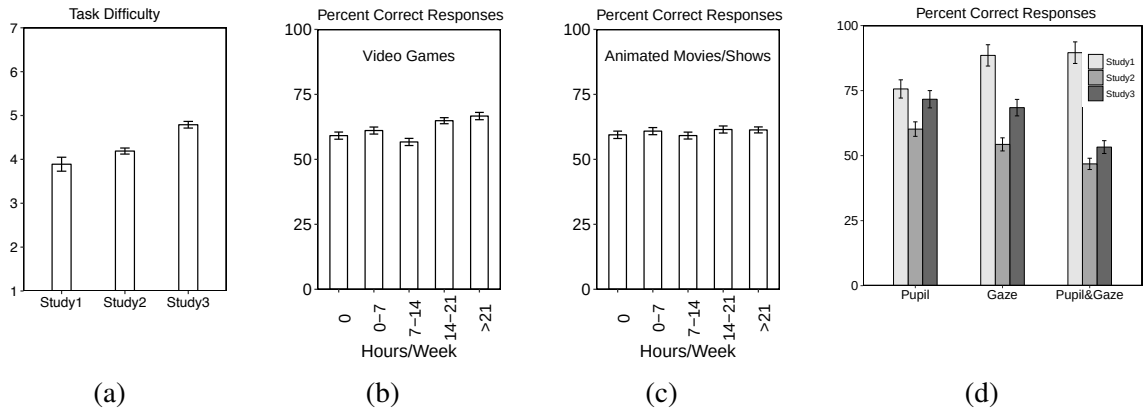

Figure 10: (a) Means and standard errors of difficulty score for Study 1, Study 2 and Study 3. (b) Means and standard errors of Percent Correct Responses for all the survey respondents in the three studies broken down by exposure to video games. (c) Means and standard errors of Percent Correct Responses for all the survey respondents in the three studies broken down by exposure to animated movies/shows. (d) Means and standard errors of Percent Correct Responses for survey respondents in Study 1, Study 2 and Study 3 broken down by their self-reported cues.

7.2 Recognition Accuracy by Exposure Level

Exposure level of survey respondents varied in our three studies. The majority of respondents spent 0-7 hours/week on video games (54.6%) as well as animated movies and TV shows (82.5%). We analyzed Percentage Correct Responses with respect to survey respondents' exposure to computer generated content to understand how this exposure might impact their judgment. We computed Percentage Correct Response as the number of stimuli where respondents identified the scared avatar as the one whose pupils were a replay of the original recording, over the total number of stimuli across all the studies. The breakdown of the average percentage of correct responses by exposure level is shown in Figure 10 (b), (c). We did not find any significant effect for exposure level.

7.3 Self-reported Cues

Survey respondents gave a free-form answer to the question "What cue(s) did you use to say whether the person is watching a scary movie?" We coded the responses as referring to one of three cues: gaze, pupil, blink. For example, if the response was "their eyes darting quickly" then this response was coded as 'gaze'. Example responses coded as 'pupil' include "pupils size increasing" and "I assume the pupil tells it all", while responses coded as 'blink' include "how much they are blinking". A response such as "Eye movements and eye pupil size change" was coded as both 'gaze' and 'pupil'. We then looked at a breakdown of Percent Correct Responses by the self-reported cues (Figure 10 (d)). In this figure, we do not consider blinks as they appear infrequently compared to gaze and pupil. Interestingly, participants thought that they used pupil and gaze in all three studies to make their judgements. In both Study 2 and Study 3, gaze was identical for the two conditions and could not have been a distinguishing cue. The known unreliability of self-reports is why we conducted a cascaded design three-part study, where we started with multiple cues present and systematically narrowed them down.

8 Discussion

In Study 1, we compared two conditions: in the first condition, the avatar's eye movements including pupil dilation, gaze and blinks matched the eye motions from people watching a scary movie, while in the second condition, the avatar's eye movements match eye motions from

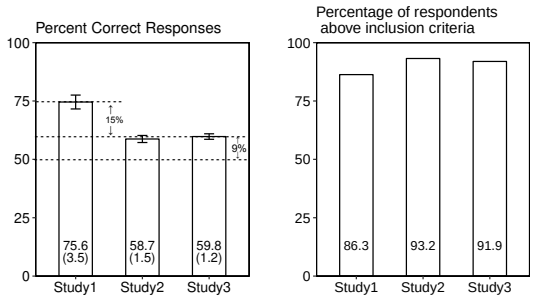

Figure 11: Left: Means and standard errors of Percent Correct Responses for Study 1, Study 2 and Study 3. Right: Percentage of respondents above consistency criteria for Study 1, Study 2 and Study 3.

the same people watching only light intensity changes. In Study 2 and Study 3, we presented experiments to evaluate if survey respondents could identify the avatar watching a scary movie when the avatar's pupil was the only factor changed (gaze direction and blinks were controlled). When we consider the three studies collectively (Figure 11), we see that the metric Percent Correct Responses is highest for Study 1, and lower for Study 2 and Study 3 (though significantly above chance in all three cases). This trend is consistent with the increasing difficulty of the task. In Study 3, the only cue that is different between the two avatars is the variation in pupil size. This cue accounts for being 9% above chance. In Study 2, two cues are different between the two avatars: average pupil size, and pupil size variation. In Study 1, the eyes of the two avatars differ in pupil, gaze, and blinks. Our findings suggest that gaze and blinks taken together add almost twice as much information as pupil size variation (15% on top of the 9% above chance). An interesting follow up experiment would be to investigate gaze and blinks separately to see if they contribute equally to this difference. In studies 2 and 3, the source of gaze and blink data is from the scary condition. It would also be interesting to test other combinations of gaze, blinks and pupil size to see if it would affect viewers judgement. We did not find any significant effect in overall correctness with respect to exposure level to animated movies or video games, suggesting that these cues are applicable to people regardless of their experience.

Our main conclusion for content creators of animated movies, video games, and virtual experiences is that getting the character's pupil motions right can help when conveying the character's emotions and bring the character to life. In all our studies, the avatar was rendered in close-up: our intention with rendering at this scale was to determine detectability. If survey respondents could not perceive pupil diameter changes even at this scale, they would definitely not do so in mid-range shots. Our findings indicate that respondents are sensitive to pupil size variation thus holds for close up views, such as zooming in to a character's face, a type of shot quite common for emotional scenes in games and movies. In film-making, such shots can capture the emotions of the characters and signal an important sensory moment in a scene. For video games, developers can use close-up shots in cut scenes, to introduce stories and conversations between characters and set the mood. Furthermore, in video games, conversational agents, and in social VR, the user can get very close to the virtual characters, and at least a head-shot distance is common.

It would also be interesting to consider different emotions and more people in future studies and to generalize the pupil effect to more settings and scenarios. We rendered out eye motions of people whose pupil diameters changed more than 1.5mm in the three seconds following the zombie attack (ID: *s004*, *s009* and *s015* in our dataset). Because this is the first study of its kind, we made this choice to detect if the effect exists. We adopted two alternative forced choice questions in the survey for the same reason. However, this paradigm did not capture the magnitude of the signal or if participants prefer another choice beyond the fear emotion. Future studies could examine the effects of pupil dynamics in other emotional states and scenarios, as well as with a larger number of participants. It would also be interesting to expand these studies to include eye motions from more people, avatars that have dark eyes, and a broader range of age groups. In Study 3, we compared eye motions from Condition Scary to Condition Procedural generated from pink noise. As future work, it would be interesting to compare Condition Scary with other baseline conditions, for

example, using the scaled pupils from Condition Neutral. For virtual avatars, it may be interesting to consider models of exaggeration and stylization. With such models it may even become feasible to directly generate eyes from an audio recording or a script. We additionally anticipate future studies that investigate the relative perceivability of pupil based cues when facial expression and dialogue are present and to test for interaction effects.

Acknowledgements

We thank Dr. Aaron Molstad for help with statistical analysis. This work was partially funded by the National Science Foundation Awards #1566481 and #1423189.

References

[1] Darryn King. Building a better digital eye. *The New Yorker*, Feb 2017.

[2] Sooha Park Lee, Jeremy B Badler, and Norman I Badler. Eyes alive. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 637–644. ACM, 2002.

[3] Sean Andrist, Tomislav Pejsa, Bilge Mutlu, and Michael Gleicher. A head-eye coordination model for animating gaze shifts of virtual characters. In *Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction*, pages 1–6. ACM, 2012.

[4] Kerstin Ruhland, Sean Andrist, Jeremy Badler, Christopher Peters, Norman Badler, Michael Gleicher, Bilge Mutlu, and Rachel McDonnell. Look me in the eyes: A survey of eye and gaze animation for virtual agents and artificial systems. In Sylvain Lefebvre and Michela Spagnuolo, editors, *Eurographics 2014 - State of the Art Reports*, pages 69–91. The Eurographics Association, 2014.

[5] Justus Thies, Michael Zollhöfer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. FaceVR: Real-time gaze-aware facial reenactment in virtual reality. *ACM Transactions on Graphics (ToG)*, 37(2), jun 2018.

[6] Pascal Bérard, Derek Bradley, Maurizio Nitti, Thabo Beeler, and Markus H Gross. High-quality capture of eyes. *ACM Transactions on Graphics (TOG)*, 33(6):223–1, 2014.

[7] Pascal Bérard, Derek Bradley, Markus Gross, and Thabo Beeler. Lightweight eye capture using a parametric model. *ACM Transactions on Graphics (TOG)*, 35(4):117, 2016.

[8] Kerstin Ruhland, Christopher E Peters, Sean Andrist, Jeremy B Badler, Norman I Badler, Michael Gleicher, Bilge Mutlu, and Rachel McDonnell. A review of eye gaze in virtual agents, social robotics and HCI: Behaviour generation, user interaction and perception. In *Computer Graphics Forum*, volume 34, pages 299–326. Wiley Online Library, 2015.

[9] Sang Hoon Yeo, Martin Lesmana, Debanga R Neog, and Dinesh K Pai. Eyecatch: simulating visuomotor coordination for object interception. *ACM Transactions on Graphics (TOG)*, 31(4):42, 2012.

[10] Laura C Trutoiu, Elizabeth J Carter, Iain Matthews, and Jessica K Hodgins. Modeling and animating eye blinks. *ACM Transactions on Applied Perception (TAP)*, 8(3):17, 2011.

[11] Amit Bermano, Thabo Beeler, Yeara Kozlov, Derek Bradley, Bernd Bickel, and Markus Gross. Detailed spatio-temporal reconstruction of eyelids. *ACM Transactions on Graphics (TOG)*, 34(4):44, 2015.

[12] Sophie Jörg, Andrew Duchowski, Krzysztof Krejtz, and Anna Niedzielska. Perceptual adjustment of eyeball and pupil diameter jitter amplitudes for virtual characters. *ACM Transactions on Applied Perception*, 15, October 2018.

[13] Krzysztof Krejtz, Andrew Duchowski, Heng Zhou, Sophie Jörg, and Anna

Niedzielska. Perceptual evaluation of synthetic gaze jitter. *Computer Animation and Virtual Worlds*, 29:e1745, 2018.

[14] Vitor F Pamplona, Manuel M Oliveira, and Gladimir VG Baranoski. Photorealistic models for pupil light reflex and iridal pattern deformation. *ACM Transactions on Graphics (TOG)*, 28(4):106, 2009.

[15] Margaret M Bradley, Laura Miccoli, Miguel A Escrig, and Peter J Lang. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4):602–607, 2008.

[16] Margaret M Bradley. Natural selective attention: Orienting and emotion. *Psychophysiology*, 46(1):1–11, 2009.

[17] Sandra P Marshall. The index of cognitive activity: Measuring cognitive workload. In *Proceedings of the 2002 IEEE 7th Conference on Human factors and power plants*, pages 7–7, 2002.

[18] Andrew T. Duchowski, Krzysztof Krejtz, Izabela Krejtz, Cezary Biele, Anna Niedzielska, Peter Kiefer, Martin Raubal, and Ioannis Giannopoulos. The index of pupillary activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 282:1–282:13. ACM, 2018.

[19] Pallavi Raiturkar, Andrea Kleinsmith, Andreas Keil, Arunava Banerjee, and Eakta Jain. Decoupling light reflex from pupillary dilation to measure emotional arousal in videos. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '16, pages 89–96, 2016.

[20] Harish Katti, Karthik Yadati, Mohan Kankanhalli, and Chua Tat-Seng. Affective video summarization and story board generation using pupillary dilation and eye gaze. In *2011 IEEE International Symposium on Multimedia (ISM)*, pages 319–326. IEEE, 2011.

[21] Congyi Wang, Fuhao Shi, Shihong Xia, and Jinxiang Chai. Realtime 3D eye gaze animation using a single RGB camera. *ACM Transactions on Graphics (TOG)*, 35(4):118, 2016.

[22] Andrew T. Duchowski, Sophie Jörg, Tyler N. Allen, Ioannis Giannopoulos, and Krzysztof Krejtz. Eye movement synthesis. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, ETRA '16, pages 147–154, 2016.

[23] Jessica Hodgins, Sophie Jörg, Carol O'Sullivan, Sang Il Park, and Moshe Mahler. The saliency of anomalies in animated human characters. *ACM Transactions on Applied Perception (TAP)*, 7(4):22:1–22:14, July 2010.

[24] Angela Tinwell, Mark Grimshaw, Debbie Abdel Nabi, and Andrew Williams. Facial expression of emotion and perception of the uncanny valley in virtual characters. *Computers in Human Behavior*, 27(2):741–749, 2011.

[25] Andrew MacQuarrie and Anthony Steed. Perception of volumetric characters' eye-gaze direction in head-mounted displays. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 645–654, 2019.

[26] William Steptoe and Anthony Steed. High-fidelity avatar eye-representation. In *2008 IEEE Virtual Reality Conference (VR)*, pages 111–114. IEEE, 2008.

[27] William Steptoe, Anthony Steed, Aitor Rovira, and John Rae. Lie tracking: social presence, truth and deception in avatar-mediated telecommunication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1039–1048. ACM, 2010.

[28] Ylva Ferstl, Elena Kokkinara, and Rachel McDonnell. Facial features of non-player creatures can influence moral decisions in video games. *ACM Transactions on Applied Perception (TAP)*, 15(1):4, 2017.

[29] Mariska E Kret. Emotional expressions beyond facial muscle actions. a call for studying autonomic signals and their impact on social perception. *Frontiers in psychology*, 6:711, 2015.

[30] Timo Partala, Maria Jokiniemi, and Veikko Surakka. Pupillary responses to emotionally provocative stimuli. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, ETRA '00, pages 123–129, 2000.

[31] Timo Partala and Veikko Surakka. Pupil size variation as an indication of affective processing. *International journal of human-computer studies*, 59(1-2):185–198, 2003.

[32] Hao Chen, Arindam Dey, Mark Billinghurst, and Robert W Lindeman. Exploring pupil dilation in emotional virtual reality environments. In Robert W. Lindeman, Gerd Bruder, and Daisuke Iwai, editors, *ICAT-EGVE 2017 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*. The Eurographics Association, 2017.

[33] Kathryn E Demos, William M Kelley, Sophia L Ryan, F Caroline Davis, and PJ Whalen. Human amygdala sensitivity to the pupil size of others. *Cerebral cortex*, 18(12):2729–2734, 2008.

[34] Shiori Amemiya and Kuni Ohtomo. Effect of the observed pupil size on the amygdala of the beholders. *Social cognitive and affective neuroscience*, 7(3):332–341, 2012.

[35] Eckhard H Hess and Slobodan B Petrovich. Pupillary behavior in communication. *Nonverbal behavior and communication*, pages 327–348, 1987.

[36] ME Kret, AH Fischer, and Carsten KW De Dreu. Pupil mimicry correlates with trust in in-group partners with dilating pupils. *Psychological science*, 26(9):1401–1410, 2015.

[37] Neil A Harrison, Tania Singer, Pia Rotshtein, Ray J Dolan, and Hugo D Critchley. Pupillary contagion: Central mechanisms engaged in sadness processing. *Social cognitive and affective neuroscience*, 1(1):5–17, 2006.

[38] Hao Li, Laura Trutoiu, Kyle Olszewski, Lingyu Wei, Tristan Trutna, Pei-Lun Hsieh, Aaron Nicholls, and Chongyang Ma. Facial performance sensing head-mounted display. *ACM Transactions on Graphics (TOG)*, 34(4):47, 2015.

[39] Katsuhiro Suzuki, Fumihiko Nakamura, Jiu Otsuka, Katsutoshi Masai, Yuta Itoh, Yuta Sugiura, and Maki Sugimoto. Recognition and mapping of facial expressions to avatar by embedded photo reflective sensors in head mounted display. In *2017 IEEE Virtual Reality (VR)*, pages 177–185, 2017.

[40] FOVE. Home - FOVE eye tracking virtual reality headset. https://fove-inc.com/product/, Retrieved Feb 12, 2021.

[41] Tobii. Tobii eye tracking: the next natural step in VR. https://vr.tobii.com/, Retrieved Feb 12, 2021.

[42] Brendan John, Pallavi Raiturkar, Arunava Banerjee, and Eakta Jain. An evaluation of pupillary light response models for 2D screens and virtual reality HMDs. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST)*, 2018.

[43] Etienne Combrisson and Karim Jerbi. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of neuroscience methods*, 250:126–136, 2015.

[44] Jacob Cohen. *Statistical power analysis for the behavioral sciences*. Routledge, 2013.

[45] Andrew Duchowski, Sophie Jörg, Aubrey Lawson, Takumi Bolte, Lech Świrski, and Krzysztof Krejtz. Eye movement synthesis with 1/f pink noise. In *Proceedings of the*

8th ACM SIGGRAPH Conference on Motion in Games, pages 47–56. ACM, 2015.

[46] Sophie Jörg, Andrew T. Duchowski, Krzysztof Krejtz, and Anna Niedzielska. Perceptual comparison of procedural and data-driven eye motion jitter. In *ACM Symposium on Applied Perception 2019*, SAP '19, 2019.