

Eye Movement Synthesis

Andrew T. Duchowski*, Sophie Jörg*, Tyler N. Allen* Ioannis Giannopoulos†
Clemson University ETH Zürich
Krzysztof Krejtz‡
University of Social Sciences and Humanities, and
National Information Processing Institute



Figure 1: Select rendered animation frames of left eye performing synthetic eye movement calibration including partial and full blinks.

Abstract

A convolution-filtering technique is introduced for the synthesis of eye gaze data. Its purpose is to produce, in a controlled manner, a synthetic stream of raw gaze position coordinates, suitable for: (1) testing event detection filters, and (2) rendering synthetic eye movement animations for testing eye tracking gaze estimation algorithms. Synthetic gaze data is parameterized by sampling rate, microsaccadic jitter, and simulated measurement error. Sampled synthetic gaze data is compared against real data captured by an eye tracker showing similar signal characteristics.

Keywords: eye movement synthesis, signal processing

Concepts: •Computing methodologies → Procedural animation; Model verification and validation;

1 Introduction & Background

The recorded eye movement signal is well understood from the point of view of analysis, but surprisingly few advances have appeared regarding its synthesis. Most analytical approaches use gaze data filtering, e.g., machine learning [Samadi and Cooke 2014], and/or processing for the detection of specific events [Ouzts and Duchowski 2012] but they rarely, if at all, mention signal synthesis. Signal processing approaches have been used for synthesis. Yeo et al.’s [2012] *Eyecatch* simulation uses the Kalman filter to produce gaze and focuses primarily on saccades and smooth pursuits. Microsaccades were not modeled. As noted by Yeo et al., simulated gaze behavior looked qualitatively similar to gaze data captured by an eye tracker, but comparison of synthesized trajectory plots showed absence of gaze jitter evident in the raw data.

*e-mail: {duchowski,sjoerg,tnallen}@clemson.edu

†e-mail: igiannopoulos@ethz.ch

‡e-mail: kkrejtz@swps.edu.pl

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 to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org. © 2016 ACM.

ETRA '16., March 14-17, 2016, Charleston, SC, USA

ISBN: 978-1-4503-4125-7/16/03

DOI: <http://dx.doi.org/10.1145/2857491.2857528>

Data-driven techniques and models based on statistical analysis of recorded movements have also been proposed. Ma and Deng [2009] described a model of gaze driven by head direction/rotation, which gave an elegant way of generating head and eye movements, although their gaze-head coupling model seemed contrary to physiology: head motion appeared to trigger eye motion. Instead, because the eyes are mechanically “cheaper” and faster to rotate, the brain usually triggers head motion when the eyes exceed a certain saccadic amplitude threshold (about 30°; see Murphy and Duchowski [2002] for a short introductory note on this topic). Nevertheless, Ma and Deng [2009] and then later Peters and Qureshi [2010] both provide useful models of gaze/head coupling with a good “linkage” between gaze and head vectors.

Le et al. [2012] provide a fully automated framework to generate realistic head motion, eye gaze, and eyelid motion based on live (or recorded) speech. To synthesize eye gaze, they first nonlinearly transform recorded gaze, speech features, and head motion to a high-dimensional feature space. Their data-driven gaze synthesis model is essentially nonlinear, but it is mainly concerned with modeling unperturbed gaze direction.

Similarly, Duchowski and Jörg [2015] model realistic eye movement rotations, focusing on adherence to Donders’ and Listing’s laws [Tweed et al. 1990]. However, their model is essentially limited to controlling torsional rotations of the eyeball. They provide a quaternion model of oculomotor rotation mechanics, but they do not discuss how their synthetic eyes can be rotated automatically.

Most of the above gaze synthesis models originate from computer graphics research (e.g., see also Wood et al. [2015]), where gaze shifts and rotations of the head are modeled to redirect gaze toward a specific target. Animating gaze shifts for virtual humans often involves the use of parametric models of human gaze behavior [Andrist et al. 2012; Pejsa et al. 2013]. While these models enable virtual humans to perform natural and accurate gaze shifts, signal characteristics, and in particular noise, are rarely addressed. Noise, however, although a nuisance from a signal processing perspective, is a key component of natural eye movements, whether it is present due to underlying neuronal properties (e.g., microsaccades), or due to eye tracker measurement noise. For animating virtual characters, the eye movement signal may even be exaggerated or stylized (i.e., *expressive*). The purpose of these models (e.g., see Vidal et al. [2015]) is to render believable characters and not necessarily to generate a signal as input to test signal analysis techniques.

The dynamics of rendered eyes have received little attention since Lee et al.’s [2002] *Eyes Alive* model which focused largely on saccadic eye movements, implementing the well-known saccadic *main*

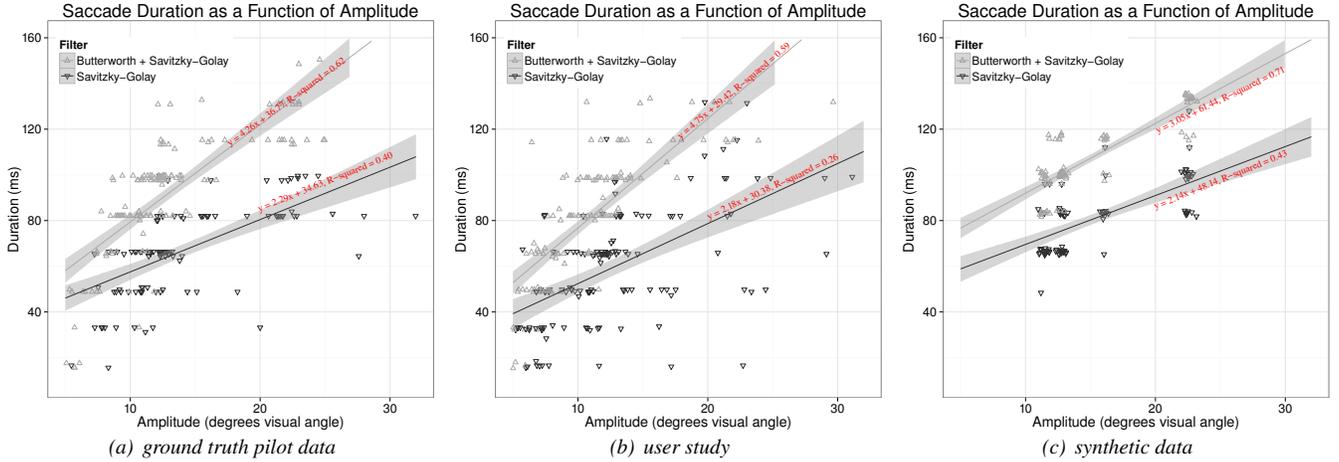


Figure 2: Saccade duration as a function of amplitude with two filtering options for data from two user studies and synthetic model.

sequence [Bahill et al. 1975]. According to Ruhland et al.’s [2014], state-of-the-art report on eyes, gaze animation largely focuses on modeling rapid saccadic shifts, smooth pursuits, binocular rotations implicated in vergence, and the coupling of eye and head rotations.

Campbell et al. [2014] propose a unified Bayesian model for frame-by-frame (rather than fixation-by-fixation) eye movement on an estimated saliency map that does not heavily rely on preprocessing of data into fixations and saccades.

The gaze synthesis models reviewed above are generally concerned with gaze direction and do not specifically address the resultant gaze signal properties. Our procedural model of gaze motion builds on Duchowski et al.’s [2015] “bottom-up” model of gaze point movement, referred to as a “look point” in space projected onto a 2D plane in front of the eye. Our simulation differs by separating out $1/f$ pink noise modeling microsaccadic jitter from eye tracker noise, modeled by white (Gaussian or normal) noise. The resulting sequence of artificial raw gaze data is used in two ways: testing event detection filters, and rendering a video of a virtual eye used to test eye-tracking algorithms. We find that the Savitzky-Golay filter is adequate for implementing velocity-based saccade detection.

2 Pilot Test for Initial Data Collection

To bootstrap the simulation, we collected data from a pilot eye tracker calibration to 11 individuals (4 male, 7 female). Conditions (e.g., apparatus, procedures) were similar to data collected later as a means of comparison with simulated data (see §4 below).

Data Filtering (Event Detection). Although Ouzts and Duchowski [2012] advocate the use of a combination of smoothing (Butterworth, or BW) and differentiation (Savitzky-Golay, or SG) filters to implement velocity-based (I-VT [Salvucci and Goldberg 2000]) event detection, we found that the inclusion of the Butterworth filter over-smooths the data resulting in outcome measures that do not match expected normal limits, i.e., the *main sequence*, relating saccadic amplitude (θ) and duration (Δt),

$$\Delta t = 2.2\theta + 21 \quad (\text{milliseconds}) \quad (1)$$

for saccadic amplitudes up to about 20° [Bahill et al. 1975; Knox 2001]. The use of the BW filter leads to overestimation of mean saccade duration, giving a main sequence of $\Delta t = 4.39\theta + 35$.

Following Nyström and Holmqvist [2010], we compared the above event detection results to the use of just the SG filter. The Savitzky-Golay [1964] filter fits a polynomial curve via least squares minimization prior to calculation of the curve’s s^{th} derivative (e.g., 1^{st} derivative ($s=1$) for velocity estimation) [Gorry 1990], hence it effects its own smoothing step prior to differentiation. This makes the use of the Butterworth filter somewhat redundant. The only reason for using the Butterworth filter could be for finer control of noise suppression that is tunable to the sampling frequency of the data. The SG filter does not provide an intuitive way of specifying sampling rate while the Butterworth filter does.

Figure 2(a) shows a comparison of the pilot calibration data captured from human participants as processed by the Butterworth and SG filter combination to processing by just the SG filter. Event detection with just the SG filter yields a main sequence equation of $\Delta t = 2.29\theta + 35$ which is inline with expectation.

The Butterworth filter used was a 2^{nd} degree filter with parameters of 60 Hz sampling rate and 6.15 Hz cutoff frequency. The Savitzky-Golay filter used was a 3^{rd} degree filter of width 5. Velocity threshold used in both cases was $40^\circ/s$.

Data Cleaning. Accuracy of raw pilot data was measured at 1.07° with precision (standard deviation) of 0.16° . Due to technical issues with the eye tracker, some time segments are missing from the data. The majority of these gaps are 300–400 ms in duration. When these gaps are present during a saccade, the duration of the gap is included as part of the saccade time. This causes excessively long saccades to be detected. For example, a saccade of 84,046 ms was produced. We removed this point from the data and all remaining saccades exceeding three standard deviations of the mean (133.73 ms). Figure 3(a) shows fixations produced by velocity-based filtering. The number of fixations appears reasonable suggesting the suitability of the Savitzky-Golay filter for implementation of the I-VT event detection algorithm. Given these results, we felt comfortable in using the SG filter for processing synthetic data and trusting its estimate of saccades and fixations.

3 Eye Movement Modeling

To generate a stream of synthetic gaze points resembling captured data $\mathbf{p}_t = (x_t, y_t)$, a reasonable strategy is to guide synthetic gaze to a sequence of known points, i.e., a grid of points that is used to

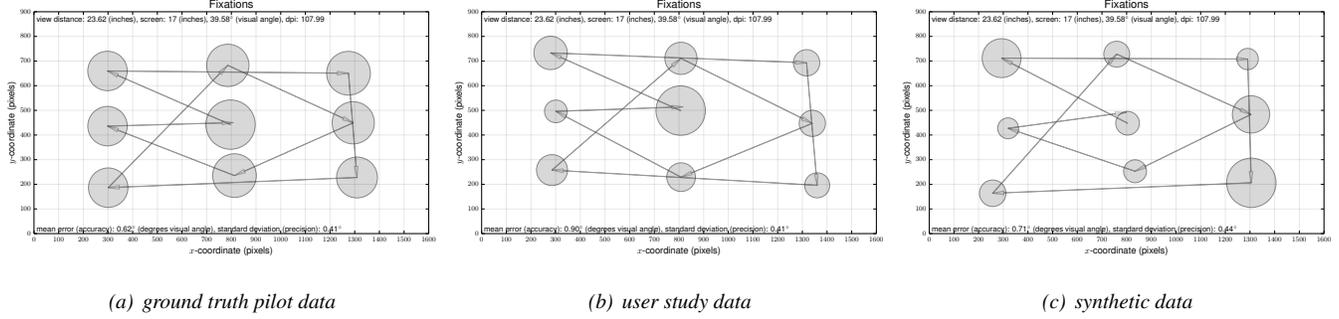


Figure 3: Representative fixations following event detection with the Savitzky-Golay filter, showing suitable I-VT event detection results. Fixation points (c) were obtained from synthetic raw data shown in Figure 4(b); note that fixation durations are randomly perturbed.

calibrate the eye tracker to human viewers. Given such a sequence (e.g., see Figure 4), several important requirements arise, namely:

1. a model of the spatio-temporal fixation perturbation;
2. a model of saccadic velocity (i.e., position and duration); and
3. control of the simulation timestep and sampling rates.

A model of spatio-temporal perturbation of gaze points at a fixation can be effected through simulation of microsaccadic jitter [Duchowski et al. 2015]. Saccadic velocity can be approximated by the main sequence (1), or it can be obtained empirically from observed data, if available. Assuming a straightforward simulation of a dynamical system involving Euler integration at each timestep, the timestep should be dissociated from the simulation’s sampling frequency. Doing so allows the simulation to produce data at different sampling rates, thus modeling various eye tracking equipment. Each of these simulation components is detailed below.

3.1 Modeling Fixations

Duchowski et al. [2015] model microsaccades by a normal distribution of $\mathcal{N}(\mu=0, \sigma=12/60)$ (arcmin) for each of the x - and y -coordinate offsets to the fixation coordinate during simulation (setting $\sigma=0$ yields no jitter during fixation), giving a sequence of points situated at the gaze coordinates given in Figure 4(a). Modeling saccadic jitter by the normal distribution yields white noise perturbation. To model microsaccades as $1/f$ pink noise perturbation, output of the white noise distribution is fed through a digital pink noise filter [Hollos and Hollos 2014]. The pink noise filter $\mathcal{P}(\alpha, \omega_0)$ is specified by two parameters: α and ω_0 , where $1/f^\alpha$ describes the pink noise power spectral distribution and ω_0 the filter’s unity gain frequency (or more simply its gain).

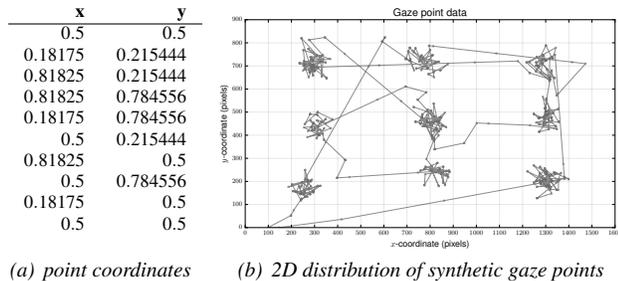


Figure 4: Generation of a sequence of synthetic gaze points.

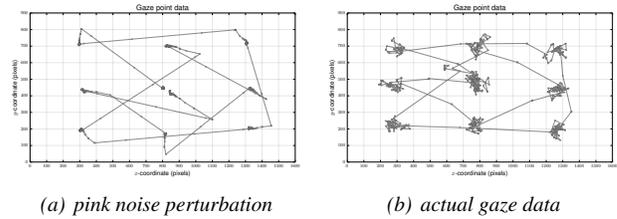


Figure 5: Synthetic pink noise perturbation and actual data.

Setting $\alpha=1$ produces $1/f$ noise, which has been observed as characteristic of pulse trains of nerve cells belonging to various brain structures [Usher et al. 1995]. Setting $\alpha=0$ produces white, uncorrelated noise, with a flat power spectral distribution, likely a poor choice for modeling biological motion such as microsaccades. We set our pink noise filter with $\alpha=0.8$ and $\omega_0=0.85$ as these are close to those used by Duchowski et al. [2015] ($\alpha=0.6$, $\omega_0=0.85$) for producing fairly realistic synthetic eye movement animations.

Synthetic data perturbed solely by pink noise produces pleasing animations of the eye, however, pink noise alone cannot account for all the noise seen in typical captured data. Figure 5 shows data perturbed solely by pink noise along with a representative (raw) gaze plot of captured data. To make synthetic gaze data useful for testing event detection filters, further signal perturbation is required.

To produce synthetic (raw) data (see Figure 4(b)), we add to the pink noise jittered fixations a perturbation modeled by white noise $\mathcal{N}(0, \sigma=1.07)$, using the average measured accuracy of the pilot raw data (see above), giving the complete fixation model as

$$\mathbf{p}_{t+h} = \mathbf{p}_t + \mathcal{P}(\alpha, \omega_0) + \mathcal{N}(0, \sigma=1.07) \quad (2)$$

where h is the simulation time step.

3.2 Modeling Saccade Acceleration, Velocity, Position

To effect movement between fixation points, a model of saccades is required, specifying both movement and duration of the gaze point. We start with an approximation to a force-time function assumed by a symmetric-impulse variability model [Abrams et al. 1989]. This function, qualitatively similar to symmetric limb-movement trajectories, describes an acceleration profile that rises to a maximum, returns to zero about halfway through the movement, and then is followed by an almost mirror-image deceleration phase.

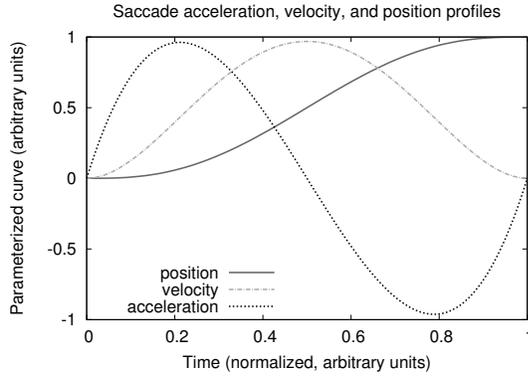


Figure 6: Parametric saccade position model derived from an idealized model of saccadic force-time function assumed by Abrams et al.’s [1989] symmetric-impulse variability model: scaled position $60H(s)$, velocity $31\dot{H}(s)$, and acceleration $10\ddot{H}(s)$.

To model a symmetric acceleration function, we can choose a combination of Hermite blending functions $h_{11}(s)$ and $h_{10}(s)$, so that $\ddot{H}(s) = h_{10}(s) + h_{11}(s)$ where $h_{10}(s) = s^3 - 2s^2 + s$, $h_{11}(s) = s^3 - s^2$, $s \in [0:1]$, and $\ddot{H}(s)$ is acceleration of the gaze point over normalized time interval $s \in [0:1]$. Integrating acceleration produces velocity, $\dot{H}(s) = \frac{1}{2}s^4 - s^3 + \frac{1}{2}s^2$ which when integrated once more produces position $H(s) = \frac{1}{10}s^5 - \frac{1}{4}s^4 + \frac{1}{6}s^3$ on the normalized interval $s \in [0:1]$ (see Figure 6).

Given an equation for position over a normalized time window ($s \in [0:1]$), we can now stretch this time window at will to any given length $s = t/\Delta t$. Because the distance between gaze target points is known *a priori*, we can use these distances (pixel distances converted to amplitudes in degrees visual angle) as input to the main sequence to obtain saccade length.

Assuming data collected from eye tracker calibration would not deviate greatly from the main sequence found in the literature [Bahill et al. 1975; Knox 2001] (actual data does not quite fit this model and differs slightly in its slope and y-intercept), we set the expected saccade duration to that given by (1) but augmented with a 10° targeting error. We also add in a slight temporal perturbation to the predicted saccade duration, based on empirical observations. Saccade duration is thus modeled as

$$\Delta t = 2.2\mathcal{N}(\theta, \sigma = 10^\circ) + 21 + \mathcal{N}(0, 0.01) \quad (\text{milliseconds}) \quad (3)$$

3.3 Running the Simulation

When running the simulation, it is important to keep the simulation time step (h) small, e.g., $h = 0.0001$. When about to execute a saccade, set the saccade clock $t = 0$, then while $t < \Delta t$ perform the following simulation steps:

1. $s = t/\Delta t$ (scale interpolant to time window)
2. $\mathbf{p}_t = \mathbf{C}_{i-1} + H(s)\mathbf{C}_i$ (advance position)
3. $t = t + h$ (advance time by the time step h)

where \mathbf{C}_i denotes the i^{th} 2D calibration point coordinates (see Figure 4(a)) and \mathbf{p}_t is the saccade position, both in vector form.

Setting time step h to an arbitrarily low value allows dissociation of the simulation clock from the sampling rate. We can thus sample the synthetic eye tracking data at arbitrary sampling periods, e.g., $d = 1$, $d = 16$, or $d = 33$ ms for sampling rates of 1000 Hz, 60 Hz, or 30 Hz, respectively. Note that sampling rates can be made very precise, generally coincident with the computer’s clock rate.

Unfortunately, eye trackers’ sampling rates are not precise, or rather, the data obtained from the eye tracker shows non-uniform sampling periods between raw gaze points (x, y, t) , most likely due to competing processes on the computer used to run the eye tracking software and/or due to network latencies.

Using the same empirical pilot data collected from calibration of an eye tracker to 11 individuals (see above), we found that the mean sampling duration was 16.41 ms with standard deviation of 1.32 ms. These descriptive statistics were obtained after removing all samples recorded with a reported sampling period greater than 1 second and all data registering at a sampling period of the mean plus three standard deviations (8.09 ms). We considered these anomalous data samples. Based on these observations, we suggest to model the sampling period by $\mathcal{N}(0, \sigma = 0.5)$ milliseconds.

3.4 Summary: Listing the Sources of Variation

To recount, the stochastic model of eye movements is based on infusion of probabilistic noise at various points in the simulation:

- fixation durations, modeled in this instance by $\mathcal{N}(1.081, \sigma = 2.9016)$ (seconds), the average and standard deviation from our pilot data,
- microsaccadic fixation jitter, modeled by pink noise $\mathcal{P}(\alpha = .8, \omega_0 = .85)$ (degrees visual angle), based on Duchowski et al.’s [2015] simulation,
- eye tracker noise, applied following fixation jitter, $\mathcal{N}(0, \sigma_x = 0.022)$ and $\mathcal{N}(0, \sigma_y = 0.036)$ in normalized screen coordinates in each of x- and y-directions, the calculated average accuracy of the x and y components, equivalent to $\mathcal{N}(0, \sigma_x = 0.78^\circ)$, $\mathcal{N}(0, \sigma_y = 0.74^\circ)$ and $\mathcal{N}(0, \sigma = 1.07^\circ)$,
- saccade durations, modeled by (3), and
- sampling period $\mathcal{N}(1, 000/\mathcal{F}, \sigma = 0.5)$ (milliseconds), with \mathcal{F} the sampling frequency (Hz).

For rendering purposes, eye tracker noise is removed, and the eye movement data stream is appended with

- blink duration, modeled as $\mathcal{N}(120, \sigma = 70)$ (milliseconds),
- pupil unrest, modeled by pink noise $\mathcal{P}(\alpha = .8, \omega_0 = .16)$ (relative diameter).

4 Comparison with Empirical Data

We evaluate our synthetic eye movement data stream in two ways. First, we compare our synthetic data to data captured by an eye tracker during calibration (using the same sequence of calibration points). Second, we compare real and synthetic scanpaths directly via cross spectral power analysis (see below). We also produce synthetic animations of the eye using Świrski and Dodgson’s [2014]¹ realistic renderer, which uses a Blender 3D model of the eye and head and a physically correct rendering technique.

Stimulus. A custom nine-point calibration was used with coordinates given in Figure 4(a). The points appeared one-by-one in sequence, starting with the central point, moving up to the upper-left corner, then proceeding clockwise to the corner points, then to the central top point, continuing in a clockwise diamond sequence until finishing with the central point.

Apparatus. In both user studies data were captured by a Gazepoint GP3 eye tracker. The GP3 was controlled by a standard Windows laptop with 8 GB RAM and Intel i7 CPU. Stimuli were presented on laptop screen (17" with 1600 × 900 resolution) with the eye tracker placed under the screen (see Figure 7). Both studies were conducted

¹<http://www.cl.cam.ac.uk/research/rainbow/projects/eyerender>

in a laboratory with no experimental conditions (the purpose of the study was solely to collect gaze data from a custom calibration sequence). The eye tracker functions at a 60 Hz sampling rate with a manufacturer-reported accuracy of $0.5\text{-}1^\circ$ with a $25 \times 11 \times 15$ cm head movement volume. Our own measurements suggest mean accuracy of the eye tracker at about 1.1° , which is what we used to model eye tracker noise (see above).

Participants. Data was captured from 16 employees of a research institute with no prior experience with eye tracking. Three participants were excluded due to calibration error (over 2°). The final sample used for statistical analyses consists of 13 participants (6 female and 7 male) with average age of 32.87 (SD=3.46). Average calibration error was 0.54° (SD=0.18°).

Procedure. The test procedure was prepared in PsychoPy [Peirce 2007]. Participants were instructed to follow and try to fixate the roving dot on the screen. They were first accustomed with the eye tracker and passed the eye tracker’s own 9-point calibration procedure. The custom calibration then started during which the gaze position data were collected. The data from the custom calibration sequence were then used in the statistical analyses reported below.

4.1 Veridical Data

The main goal of the statistical analysis was to compare linear regression models between saccade amplitude and duration produced by different filter combinations, i.e., either the Savitzky-Golay filter or a combination of the Butterworth and Savitzky-Golay filters.

First, we describe the statistical test of differences in saccade amplitude and duration between data produced by both filters. Second, the regression linear models of relation between saccade duration and amplitude were fit and their slopes were compared with the use of linear models. Finally, bootstrap resampling simulations were run to obtain reliable confidence intervals of intercept and slope of the regression models for comparison with the expected main sequence (1). All statistical analyses were scripted with R, the language for statistical computing.

The comparison of mean saccade amplitudes produced by the filter combinations with a Welch two sample t-test revealed statistical significance, $t(251.61) = 2.76, p < 0.001$. Saccade amplitude was significantly greater without the Butterworth filter ($M = 12.69, SE = 0.52$) than with the Butterworth filter ($M = 10.90, SE = 0.38$). Similar comparison of saccade duration showed a statistically significant

difference between both filter combinations, $t(275.2) = 7.20, p < 0.001$. The Savitzky-Golay filter produced significantly shorter saccade durations ($M = 58.02, SE = 2.23$) than the Butterworth/SG filter combination ($M = 81.23, SE = 2.33$).

The use of Butterworth/Savitzky-Golay combination results in a main sequence equation of $\Delta t = 4.75\theta + 29.42$. The use of just the SG filter yields a main sequence equation of $\Delta t = 2.18\theta + 30.38$.

To test the difference in slope and intercept of the linear relation between saccade duration and amplitude, linear modeling was used with the interaction term between predictors. Saccade duration was treated as the dependent variable with amplitude and filter used as predictors. Analysis revealed that the whole regression model was statistically significant, $F(3, 274) = 99.97, p < 0.001$. It also showed that intercepts for the two filter combinations were not significantly different ($\Delta\beta_{intercept} = -0.95, SE = 6.04, t(274) = 0.16, p = 0.88$). However, the difference in slopes ($\Delta\beta_{slope} = 2.57, SE = 0.48$) was statistically significant, $t(274) = 5.37, p < 0.001$, see Figure 2(b).

In order to test whether the coefficients of linear regression models obtained from the filtered data are similar to the main sequence, the regression models were bootstrapped using 2,000 iterations. Bootstrapping allows for estimation of accurate confidence intervals of regression coefficients based on relatively small sample sizes, as in the case of this study.

For data filtered with the Savitzky-Golay filter, the 95% confidence interval for the intercept is [20.56, 44.60] while for slope [1.00, 3.05]. For data filtered with the Butterworth/Savitzky-Golay combination the confidence interval for the intercept is [19.52, 40.10] and for slope it is [3.92, 5.63]. The main sequence slope (2.2) fits within the confidence interval produced by data filtered with just the Savitzky-Golay filter. The main sequence intercept (21) fits within the confidence intervals obtained from data processed by both filter combinations.

Summing up, veridical results of the user study showed several significant differences between data produced by the Savitzky-Golay and the Butterworth/Savitzky-Golay filter combinations. Data analyzed with just the SG filter produced significantly higher saccade amplitudes and lower durations than the BW/SG combination. The linear relation of saccade amplitudes and durations produced by the SG filter have significantly shallower slope than the regression slope obtained from the BW/SG combination. However, they both produce similar intercepts. Finally, the bootstrap resampling procedure showed that data filtered with the Savitzky-Golay filter led to a fit regression line between saccade amplitude and duration that is similar to the original main sequence of (1) [Bahill et al. 1975].

4.2 Synthetic Data

The effect of filter combinations is similar on synthetic data as it is on real data. Figure 2(c) shows a comparison of the synthetic data as processed by the Butterworth and SG filter combination with processing by just the SG filter. The use of Butterworth filter smoothing prior to application of the SG filter overestimates mean saccade duration, resulting a main sequence equation of $\Delta t = 2.85\theta + 65$.

Event detection with just the SG filter yields a main sequence equation of $\Delta t = 2.35\theta + 42$ which is more inline with expected outcomes since the data is generated starting with main sequence equation. Statistically, the difference in mean saccadic duration resulting from the two event detection methods, as computed by a Welch two sample t-test, is significant ($p < 0.01$) The Butterworth/Savitzky-Golay combination yields a much higher estimate of mean saccadic duration ($M = 106.21, SD = 14.73$) than as estimated by just the Savitzky-Golay filter ($M = 77.66, SD = 14.13$). Filter parameters were the same throughout (see §2).

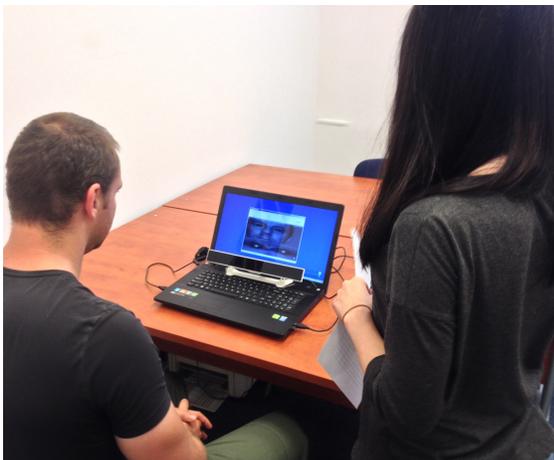


Figure 7: Experimental setup with participant and experimenter.

Table 1: Comparison of saccade durations and amplitude fits from pilot and veridical user studies to synthetic data processed with the SG or BW/SG filter combinations via regression analysis.

Filter	Statistic	Study	$\Delta\beta$	SE	<i>t</i> -test	<i>p</i> -value
SG (df=334)	intercept	pilot	-6.99	8.26	0.85	0.40
		veridical	-11.24	7.52	1.49	0.14
	slope	pilot	-0.05	0.53	0.11	0.92
		veridical	-0.17	0.48	0.35	0.73
BW/SG (df=352)	intercept	pilot	-30.13	7.67	3.93	< 0.001
		veridical	-35.20	7.59	4.64	< 0.001
	slope	pilot	1.54	0.52	2.94	< 0.01
		veridical	1.91	0.56	3.55	< 0.001

Testing whether the coefficients of linear regression models obtained from the filtered data are similar to the main sequence, bootstrapping with 2,000 iterations was applied once again. For data filtered with the Savitzky-Golay filter, the 95% confidence interval for the intercept is [34.49, 48.54] while for the slope [1.91, 2.75]. The confidence interval of the slope of data filtered with just the SG filter contains the main sequence slope value (2.20). The obtained slope and intercept of data filtered with the Butterworth/Savitzky-Golay combination produce the following 95% confidence intervals: for the intercept is [54.68, 71.80] and for the slope [2.39, 3.39]. Neither confidence interval contains the main sequence intercept or slope.

5 Results: Synthetic vs. Veridical Data

To verify whether the synthetic data sufficiently matches data obtained from both the pilot and user studies, we compared the linear relations (via slope and intercept) between saccade duration and amplitude. Linear regression analyses were conducted where amplitude was a continuous predictor, with the source of the data (synthetic, pilot, and user study) as a dichotomous predictor. The analysis was run separately for data processed by just the Savitzky-Golay filter and the Butterworth/Savitzky-Golay filter combination. In both analyses, regression lines from pilot and veridical data were compared to the regression line fit to synthetic data.

Analysis of data processed with the SG filter revealed that the regression model was statistically significant, $F(5, 334) = 46.76, p < 0.001$, meaning that the intercept ($\beta = 41.62, SE = 6.72, t(334) = 6.19, p < 0.001$) and slope ($\beta = 2.35, SE = 0.42, t(334) = 5.59, p < 0.001$) of the relation between amplitude and duration were significant, independent of the data source. What is of interest to this study is that the analyses showed that the difference between veridical data and the model fit to the synthetic data was not significant, either in terms of intercept or slope (see Table 1, c.f. Figure 2).

Contrary results were obtained via analysis of data processed with the BW/SG filter combination. The model, independent of data source, was also significant, $F(5, 352) = 131.10, p < 0.001$, showing a statistically significant relation between amplitude and duration described by intercept ($\beta = 64.62, SE = 6.69, t(352) = 9.66, p < 0.001$) and slope ($\beta = 2.85, SE = 0.44, t(352) = 6.45, p < 0.001$). However, the regression lines fit to both pilot and veridical data differed significantly in their intercepts and slopes when compared to the regression line fit to the synthetic data. For detailed statistics see Table 1 (c.f. Figure 2).

Summing up, regression analyses show that synthetic data is statistically similar to the two sets of veridical data, when processed by the Savitzky-Golay filter. In terms of saccade amplitudes and duration, synthetic data serves as a suitable match to authentic eye movement data. The addition of the Butterworth filter into the signal processing pipeline alters the slope and intercept of the linear regression, significantly deviating from veridical data.

5.1 Cross Spectral Power Analysis

Analysis via main sequence modeling compares synthetic and veridical data in terms of saccade duration and saccade amplitude. Direct comparison of eye movement data is problematic since available approaches do not necessarily consider stochastic signal properties. Probability-distance measures such as the Kullback-Leibler (KL) divergence or the Earth Mover's Distance (EMD; see for example Dempere-Marco et al. [2010]) are generally designed to compare spatial distributions (e.g., of aggregated fixations), without necessarily considering how the spatial distribution(s) came to be, i.e., characteristics of the process that created them.

Instead of comparing spatial distributions (e.g., of fixations), we need rather to inspect the underlying dynamical processes governing their creation. That is, we require spectral analysis, which considers the spectral content (i.e., the distribution of power over frequency) of a time series [Stoica and Moses 2005]. Specifically, we consider the pairwise signal coherence, the normalized Cross Spectral Density (CSD) of raw gaze data streams represented as time series in the frequency domain.

Raw data stream timestamps do not match due to imperfect sampling, e.g., while sampling at 60 Hz data is theoretically recorded every 16 ms on average, but in practice this is imprecise and may range by a few milliseconds. To temporally align the data, it must be resampled: since we cannot upsample, we downsample to a common minimum rate, i.e., 50 Hz since we expect temporal gaps of no more than 20 ms in the (raw) data streams (sampled at 60 Hz).

Raw gaze data is typically 2D with *x*- and *y*-components. For computation of the CSD, we only consider 1D time series and use only the *x*-coordinate (i.e., left-to-right gaze jitter). With raw eye movement data defined as $\mathbf{p}_t = (x_t, y_t)$, we consider only the *x*-coordinate x_t and treat it as a discrete-time signal $u = x(t)$.

The signal coherence, or normalized Cross Spectral Density (CSD),

$$C^2(\omega) = \frac{|\phi_{uv}(\omega)|^2}{\phi_{uu}(\omega)\phi_{vv}(\omega)}$$

is analogous to correlation but in this case between two time series *u* and *v*, where ϕ_{uv} is the CSD of the two series and ϕ_{uu} and ϕ_{vv} are the Power Spectral Densities (PSDs) of each. Power Spectral Density (PSD) is defined as the discrete-time Fourier transform of the signal covariance $\phi_u(\omega) = \sum_{k=-\infty}^{\infty} r(k)e^{-i\omega k}$ with $r(k)$ the autocovariance sequence of $x(t)$ defined as $r(k) = E\{x(t)x^*(t-k)\}$ with $E\{\cdot\}$ denoting the expectation operator (which averages over the ensemble of realizations), where it is assumed to depend only on the lag between two samples averaged [Stoica and Moses 2005].

Coherence is a function of frequency; to compute a single similarity metric between a pair of signals, we integrate over frequency to obtain total power (or variance in a statistical sense) $P = \frac{1}{T} \int_0^T C^2(\omega)$ where *T* is the extent of frequency components sampled ($T \in [0 : 0.25]$ Hz in this case; see Figure 9). The CSD of a signal with itself produces no variance (in the statistical sense) and hence $P = 1$, giving a convenient, normalized metric of similarity.

To test similarity of our synthetically produced data to veridical data we considered all raw data for which we had at least 500 samples that contained temporal sample gaps no greater than 20 ms: 11 raw data streams from our pilot data (reabeled PD), 12 raw data streams from our user study (reabeled US), 11 raw data streams with synthetic jitter but no eye tracker noise (SJ), and 11 raw data streams with both synthetic jitter and eye tracker noise (SN), for a total of 45 data streams, yielding $\binom{45}{2} = 990$ combinations (we also tested each stream with itself but omitted these unity results in the final statistical analyses).

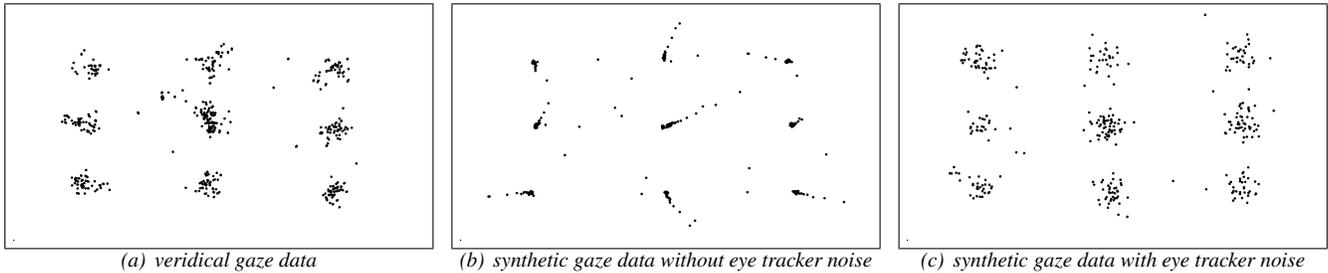


Figure 8: Gaze point (look point) composite renderings showing the entire calibration sequence on one frame. Both veridical and synthetic data augmented with noise show similarly large spatial distributions about the calibration points. Synthetic data without noise appears more realistic when looking at the eye movement animations.

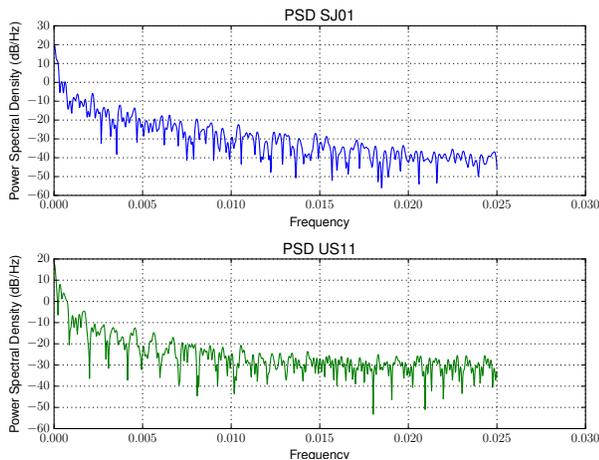


Figure 9: Power Spectral Densities (PSDs) of synthetic gaze jitter (SJ01, without eye tracker noise), above, and veridical data (US11) below. For this pairing, $P = 0.16$ suggesting very low spectral coherence between the two signals.

Extracting the pilot data stream from the complete list of pairings and using it as a blocking factor, we conducted a one-way ANOVA on power considering the type of data stream as the fixed factor at 3 levels. For the pilot data (PD), a significant effect was noted ($F(2, 371) = 22.2, p < 0.01$), with mean power significantly lower between the pilot data (PD) and jitter data (SJ) pairings. Pairwise t-tests with Bonferroni correction showed a significant difference between SJ and each of the US ($p < 0.01$) and SN ($p < 0.01$) mean power similarities. No significant difference was observed between the SN and US data streams meaning that the mean cross spectral density power between the pilot data and each of the user study and synthetic data streams (with noise) was not statistically different.

Similar results were obtained when each of the three other data streams were used as blocking factors. What the results indicate is that, on average, the stochastic properties of the synthetically generated signal are similar to veridical data, when simulated eye tracker noise is present. Omitting eye tracker noise from the simulation produces a stochastically different signal.

6 Rendering Synthetic Data

For rendering of synthetic gaze, whether for eye tracker algorithm testing as is the current objective, or alternatively for virtual character animation, the synthetic stream of eye movement positions

(x, y, t) generated by the above stochastic model can effectively be treated a series of “look points” projected onto a virtual image plane in front of the avatar or eye model. Similarly, gaze data captured by an eye tracker, often normalized, can also be used to drive the animation, although because of the inherent noise present in the signal, resulting animations are much too jittery to be believable. Figure 8 shows the composite distribution of sampled raw data: both veridical and synthetic data augmented with noise show similarly large spatial distributions about the calibration points. Meanwhile, the synthetic data rendered only with microsaccadic jitter appears more realistic when looking at the eye movement animations. To produce believable synthetic eye movements, therefore, the signal provided to the renderer must be such that fixations do not include the eye tracker noise in fixations modeled by (2).

7 Discussion & Limitations

Statistical analyses of the synthetic data regarding its resultant main sequence ($\Delta t = 2.35\theta + 42$), when processed with just the Savitzky-Golay filter, shows that the main sequence linear fit: (a) matches the main sequence intercept (1) reported in the literature (via statistical bootstrapping), and (b) matches the main sequence linear fits of both sets of veridical data. We are confident that our synthetic data can readily be used for testing event detection filters. Once measurement noise is removed from the simulation, the resultant data stream should be suitable for rendering synthetic eye movements.

Current statistical analysis is largely limited by the sampling rate (60 Hz) of the eye tracker. Following Nyquist’s theorem, we are limited in observation of saccades of 33 ms minimum duration. Clearly this will miss much shorter saccades. We will revisit our modeling and evaluation methods with a 150 Hz tracker from Gaze-point, which will allow detection of saccades of 13 ms duration.

8 Conclusion

We provide a stochastic model of gaze that is suitable for both eye movement animation rendering and for testing event detection filters. The latter utility was demonstrated by testing two filter combinations and showing that the inclusion of the Butterworth filter tends to overestimate saccade durations, if not used carefully. The former utility was demonstrated by rendering eye movements for subsequent testing of eye tracking gaze point estimation algorithms. We showed empirically that rendering data directly output by the eye tracker would be unrealistically noisy. A more reasonable eye movement synthesis model must separate measurement noise from underlying (microsaccadic) jitter associated with fixations.

Acknowledgments

The work is supported in part by US NSF Grant No. IIS-1423189. We would like to express our sincere thanks to Anna Niedzielska and Dr. Cezary Biele for their help in running the studies.

References

- ABRAMS, R. A., MEYER, D. E., AND KORNBUM, S. 1989. Speed and Accuracy of Saccadic Eye Movements: Characteristics of Impulse Variability in the Oculomotor System. *Journal of Experimental Psychology: Human Perception and Performance* 15, 3, 529–543.
- ANDRIST, S., PEJSA, T., MUTLU, B., AND GLEICHER, M. 2012. Designing effective gaze mechanisms for virtual agents. In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems*, ACM, CHI '12, 705–714.
- BAHILL, A. T., CLARK, M., AND STARK, L. 1975. The Main Sequence, A Tool for Studying Human Eye Movements. *Mathematical Biosciences* 24, 3/4, 191–204.
- CAMPBELL, D. J., CHANG, J., CHAWARSKA, K., AND SHIC, F. 2014. Saliency-based bayesian modeling of dynamic viewing of static scenes. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ACM, ETRA '14, 51–58.
- DEMPERE-MARCO, L., HU, X.-P., AND YANG, G.-Z. 2010. A Novel Framework for the Analysis of Eye Movements during Visual Search for Knowledge Gathering. *Cognitive Computation*, 1–17.
- DUCHOWSKI, A. T., AND JÖRG, S. 2015. Modeling Physiologically Plausible Eye Rotations. In *Proceedings of Computer Graphics International (Short Papers)*, CGI '15.
- DUCHOWSKI, A., JÖRG, S., LAWSON, A., BOLTE, T., ŚWIRSKI, L., AND KREJTZ, K. 2015. Eye Movement Synthesis with $1/f$ Pink Noise. In *Motion in Games (MIG)*, ACM.
- GORRY, P. A. 1990. General Least-Squares Smoothing and Differentiation by the Convolution (Savitzky-Golay) Method. *Analytical Chemistry* 62, 6, 570–573.
- HOLLOS, S., AND HOLLOS, J. R. 2014. *Creating Noise*. Exstrom Laboratories, LLC, Longmont, CO, April. ISBN: 9781887187268 (ebook), URL: <http://www.abrazol.com/books/noise/> (last accessed Jan. 2015).
- KNOX, P. C. 2001. The Parameters of Eye Movement. Lecture Notes, URL: <http://www.liv.ac.uk/~pcknox/teaching/Eymovs/params.htm> (last accessed Nov. 2012).
- LE, B., MA, X., AND DENG, Z. 2012. Live speech driven head-and-eye motion generators. *Visualization and Computer Graphics*, *IEEE Transactions on* 18, 11 (Nov), 1902–1914.
- LEE, S. P., BADLER, J. B., AND BADLER, N. I. 2002. Eyes Alive. *ACM Transactions on Graphics* 21, 3 (July), 637–644.
- MA, X., AND DENG, Z. 2009. Natural Eye Motion Synthesis by Modeling Gaze-Head Coupling. In *IEEE Virtual Reality*, 143–150.
- MURPHY, H., AND DUCHOWSKI, A. T. 2002. Perceptual Gaze Extent & Level Of Detail in VR: Looking Outside the Box. In *Conference Abstracts and Applications (Sketches & Applications)*, ACM. SIGGRAPH Annual Conference Series.
- NYSTRÖM, M., AND HOLMQVIST, K. 2010. An Adaptive Algorithm for Fixation, Saccade, and Glissade Detection in Eyetracking Data. *Behaviour Research Methods* 42, 1, 188–204.
- OUZTS, A. D., AND DUCHOWSKI, A. T. 2012. Comparison of Eye Movement Metrics Recorded at Different Sampling Rates. In *Eye Tracking Research & Applications (ETRA)*, ACM, 321–324.
- PEIRCE, J. 2007. PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods* 162, 1-2, 8–13.
- PEJSA, T., MUTLU, B., AND GLEICHER, M. 2013. Stylized and Performative Gaze for Character Animation. In *Proceedings of EuroGraphics*, EuroGraphics, I. Navazo and P. P., Eds.
- PETERS, C., AND QURESHI, A. 2010. A Head Movement Propensity Model for Animating Gaze Shifts and Blinks of Virtual Characters. *Computers & Graphics* 34, 677–687.
- RUHLAND, K., ANDRIST, S., B., B. J., PETERS, C. E., BADLER, N. I., GLEICHER, M., MUTLU, B., AND MCDONNELL, R. 2014. Look me in the eyes: A survey of eye and gaze animation for virtual agents and artificial systems. In *Computer Graphics Forum*, Lefebvre, S. and Spagnuolo, M., Ed., EuroGraphics STAR—State of the Art Report.
- SALVUCCI, D. D., AND GOLDBERG, J. H. 2000. Identifying Fixations and Saccades in Eye-Tracking Protocols. In *Eye Tracking Research & Applications (ETRA) Symposium*, ACM, 71–78.
- SAMADI, M. R. H., AND COOKE, N. 2014. EEG Signal Processing for Eye Tracking. In *Proceedings of the Signal Processing Conference (EUSIPCO)*, IEEE, 2030–2034.
- SAVITZKY, A., AND GOLAY, M. J. E. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* 36, 8, 1627–1639.
- STOICA, P., AND MOSES, R. 2005. *Spectral Analysis of Signals*. Prentice Hall, Upper Saddle River, NJ.
- ŚWIRSKI, L., AND DODGSON, N. 2014. Rendering synthetic ground truth images for eye tracker evaluation. In *Eye Tracking Research and Applications*, ACM, ETRA '14, 219–222.
- TWEED, D., CADERA, W., AND VILIS, T. 1990. Computing Three-dimensional Eye Position Quaternions and Eye Velocity From Search Coil Signals. *Vision Research* 30, 1, 97–110.
- USHER, M., STEMMLER, M., AND OLAMI, Z. 1995. Dynamic Pattern Formation Leads to $1/f$ Noise in Neural Populations. *Physical Review Letters* 74, 2, 326–330.
- VIDAL, M., BISMUTH, R., BULLING, A., AND GELLERSEN, H. 2015. The royal corgi: Exploring social gaze interaction for immersive gameplay. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM, New York, NY, CHI '15, 115–124.
- WOOD, E., BALTRUSAITIS, T., ZHANG, X., SUGANO, Y., ROBINSON, P., AND BULLING, A. 2015. Rendering of eyes for eye-shape registration and gaze estimation. *CoRR abs/1505.05916*.
- YEO, S. H., LESMANA, M., NEOG, D. R., AND PAI, D. K. 2012. Eyecatch: Simulating Visuomotor Coordination for Object Interception. *ACM Transactions on Graphics* 31, 4 (July), 42:1–42:10.