Investigating the Effects of Eye-Tracking Interpolation Methods on Model Performance of LSTM

Jesse W. Grootjen
LMU Munich
Munich Center for Machine Learning
Munich, Germany
jesse.grootjen@ifi.lmu.de

Henrike Weingärtner
LMU Munich
Munich, Germany
henrike.weingaertner@ifi.lmu.de

Sven Mayer
LMU Munich
Munich Center for Machine Learning
Munich, Germany
info@sven-mayer.com

ABSTRACT

Physiological sensing enables us to use advanced adaptive functionalities through physiological data (e.g., eye tracking) to change conditions. In this work, we investigate the impact of infilling methods on LSTM models' performance in handling missing eye tracking data, specifically during blinks and gaps in recording. We conducted experiments using recommended infilling techniques from previous work on an openly available eye tracking dataset and LSTM model structure. Our findings indicate that the infilling method significantly influences LSTM prediction accuracy. These results underscore the importance of standardized infilling approaches for enhancing the reliability and reproducibility of LSTM-based eye tracking applications on a larger scale. Future work should investigate the impact of these infilling methods in larger datasets to investigate generalizability.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**.

KEYWORDS

human computer interaction

ACM Reference Format:
Jesse W. Grootjen, Henrike Weingärtner, and Sven Mayer. 2024. Investigating the Effects of Eye-Tracking Interpolation Methods on Model Performance of LSTM. In *2024 Symposium on Eye Tracking Research and Applications (ETRA '24), June 04–07, 2024, Glasgow, United Kingdom.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3649902.3656353

1 INTRODUCTION

Nowadays, eye tracking has emerged as an additional input channel for multi-modal interactions, as evidenced by studies such as [Esteves et al. 2015; Lischke et al. 2016; Turner et al. 2014]. However, optical and infrared eye tracking data are susceptible to data loss, particularly when the eye tracker encounters challenges in estimating pupil direction, frequently occurring during human blinks. This frequent data loss poses challenges for traditional methods of comprehending user behaviors and prediction models, including intent prediction, necessitating additional pre-processing steps. Consequently, various blink detection methods and strategies to address gaps in the input data stream have been explored. Grootjen et al. [2023] underscored the absence of standardized processes to overcome challenges posed by eye blinks, introducing a significant impediment to the reproducibility and comparability of findings across diverse studies. Following this, Grootjen et al. [2024] showed the inconsistencies in reporting and the influence different use-case-specific approaches have on the internal validity of eye-tracking studies. While some interactive systems disregard input affected by missing data, this practice introduces input lag and unexpected jumps and jitters, substantially diminishing system usability [Lugrin et al. 2013; MacKenzie and Ware 1993]. As machine learning methods, such as recurrent neural networks (RNN) and long short-term memory (LSTM), gain popularity for processing and predicting interactions, their effectiveness is contingent on a consistent data input stream without gaps. Consequently, the current understanding of the impact of the different infilling methods on the quality of these LSTM and RNN models.

Diverse interactive systems leverage eye tracking to enhance functionality, encompassing applications like direct manipulation [Lischke et al. 2016; Pfeuffer and Gellersen 2016], action prediction [Zhang et al. 2022], and gestures [Drewes and Schmidt 2007; Zhang et al. 2017]. Recent advancements have seen these systems incorporating neural networks to refine traditional feature extraction methods, as demonstrated by studies such as [Aftab et al. 2020; Zhang et al. 2022]. However, the inherent challenge lies in the neural networks' limited capacity to handle missing information during blinks effectively. Consequently, prevalent strategies for dealing with data loss from eye trackers involve either excluding data with blinks, as observed in [Ekman et al. 2008; Gunawardena et al. 2019; Wang et al. 2021], or attempting to fill in the missing information, as explored by Stein et al. [2022], utilizing use-case-specific and device-specific approaches. Regrettably, these studies often neglect concerns of reproducibility and generalizability, omitting evaluations of the impact of fine-tuning, such as specific parameters for infilling methods. Moreover, blinks introduce artifacts into the retained eye tracking data, and despite their acknowledged presence in the literature [Abel et al. 1983; Collewijn et al. 1985; Epelboim and Suppes 2001], present-day systems frequently overlook addressing these artifacts, leading to a general tendency to ignore the affected input. Consequently, there is an imperative need to establish a comprehensive and consistent pre-processing approach for eye tracking data to ensure the reliability and validity of interactive systems utilizing eye tracking.

ETRA '24, June 04–07, 2024, Glasgow, United Kingdom

ACM ISBN 979-8-4007-0607-3/24/06
https://doi.org/10.1145/3649902.3656353

Evaluating the impact of these different infilling methods on openly available datasets will bring an understanding of potential generalization issues and will allow us to formalize recommendations to overcome them. Therefore, future work will know which of these infilling methods to apply without interfering with the output of their machine learning models. Thus, we have used the recommended infilling methods from Grootjen et al. [2024] on an openly available dataset from Annerer-Walcher et al. [2021]. By using their data and model structure, we can observe the impact of the different infilling methods on the accuracy of the LSTM model.

In this work, we found that linear and cubic spline interpolation over missing values and gaps in eye tracking data has a major impact on the classification accuracy of their LSTM model, to the point that overfitting occurred. This highlights the importance of further investigation into the impact of interpolation methods on the accuracy of LSTM or RNN models of different datasets to investigate generalizability.

2 RELATED WORK

First, we provide a short overview of the reasons for blinks and how blinks are used in systems for human-computer interaction (HCI). Next, we provide insight into different ways of blink detection. For the final part of our related work, we provide various ways of dealing with these blinks for machine learning methods.

2.1 Reasons for Blinks

Blinking is "a temporary closure of both eyes, involving movements of the upper and lower eyelids" [Blount 1927]. Human adults blink on average 12 times per minute, and one blink lasts roughly one-third of a second [Fatt and Weissman 2013]. Blinks protect the eye from drying out and regularly replenish the precorneal tear-film. However, there are a large variety of factors impacting the blink frequency of a human outside of these responsibilities, including but not limited to the time of day [Stern et al. 1994], the presence of air pollutants [Stern et al. 1994], monitors [Patel et al. 1991], contact lenses [Collins et al. 1989], perceptual load [Brookings et al. 1996; Tsai et al. 2007; Van Orden et al. 2000; Wolkoff et al. 2005], age [Stern et al. 1994], psychoticism [Colzato et al. 2009], and individual differences [Doughty and Naase 2006].

Various human-computer interaction (HCI) studies use blink data in interactive systems such as lie detection [Leal and Vrij 2008; Mann et al. 2002], driver fatigue detection [Bergasa et al. 2006; Hernandez-Ortega et al. 2019b], detection of mild cognitive impairment [Ladas et al. 2014], anti-face spoofing [Galbally et al. 2014; Hernandez-Ortega et al. 2019a; Pan et al. 2007], and human-computer interfaces [Acien et al. 2020] among many others. However, as the frequency of blinks is influenced by many factors, the accuracy of these interactive systems can be heavily impacted.

2.2 Dealing with Blinks

Grootjen et al. [2023] highlights the importance of consistently handling missing data as it hinders the development of effective intelligent systems, limits reproducibility, and can even lead to incorrect results. Although there are various parsers available for detecting and dealing with blinks, in Grootjen et al. [2024], the authors found that these are not always used and that the general way of dealing with blinks in eye tracking data is inconsistent. Furthermore, they compare different infilling methods for missing data and the error these methods produced in a set of artificially introduced blinks. In their work, the compared infilling methods were extracted from a literature review. They found that linear and cubic spline interpolation within the missing data produced the slightest error and that artifacts from blinks affect the eye movements 70 ms preceding and 118 ms following a blink.

2.3 Long Short-Term Memory Neural Networks

Long short-term memory (LSTM) is a deep neural network architecture that can classify time-series data. This technique benefits over traditional machine learning as it does not require domain-specific knowledge as it benefits from representation learning. This might be the reason for its rise in popularity in the physiological signal space. An example is the work of Pham [2021]; here, they used it for classifying ECG data. Moreover, it has also been gaining popularity in the eye tracking community (e.g., Bremer et al. [2023]; Hassan et al. [2022]; Palacios-Ibáñez et al. [2023]; Stein et al. [2022]).

3 METHOD

We base our analyses on the data and scripts from Annerer-Walcher et al. [2021] to evaluate the different infilling methods. We selected this dataset as, from their work, they provide both their data and model structure open-source on https://osf.io/scmry/. Their dataset consisted of binocular eye tracking data, including x and y-screen-based coordinates and pupil dilation. It contains information on conditions internal and external focus vs. verbal, numerical, and visuospatial tasks (two conditions × three tasks). They investigated how consistently different eye parameters respond to internal versus external attentional focus across the three task modalities (verbal, numerical, and visuo-spatial). They report that classifying the focus of attention worked well across participants but that generalizing it across the different tasks had proven challenging.

We use their data and model structure to evaluate the impact of the different infilling methods from peer-reviewed work. As such, we ran their scripts in 3 fold, once without changes, once where we linearly interpolated the gaps and once where the interpolation was done using a cubic spline method, without altering their scripts to preserve the validity of our work. These are the recommended infilling methods as by Grootjen et al. [2024]. Following those guidelines, we removed 70 ms preceding and 118 ms following missing data, as the blink can affect these. We used scripts containing the interpolation methods and removed the data preceding and following a blink from Grootjen et al. [2024], as these are openly available on https://eyetrackingguidelines.github.io/.

3.1 Pre-Processing the Data

To preprocess the data, we leveraged the existing scripts from Annerer-Walcher et al. [2021]. These existing scripts allowed us to read the different files that are part of the main task. They also provided training before the main task on their open-science framework repository. Even though we could not find in their code that this explicitly was excluded, we assumed it was and thus only used the files from the main task. The SMI RED250mobile system (SensoMotoric Instruments, Germany) with a temporal resolution of 250

Hz, a spatial resolution of 0.03°, and gaze position accuracy of 0.4° that they used for their experiment, writes 0 in the file whenever the eye tracker cannot find the pupil[1]. To leverage the existing scripts for interpolation from Grootjen et al. [2024], we replaced all of the "zeros" with "Not a Number's" (NaN).

3.2 Missing Data

Figure 1a presents the results of the consecutive NaN's logged. In the work of Grootjen et al. [2024], they showed that the closed eye time does not go beyond 1 second. As such, we have split the dataset into different parts once this happens, as the assumption here would be that there is missing values for other reasons than a blink. When looking at the remaining consecutive NaN's in the data we can see that this follows a similar pattern as the one visualized in Grootjen et al. [2024]; Holmqvist et al. [2011]. We use a Generalized Inverse Gaussian distribution [Perreault et al. 1999] to model the distributions of the lengths, see Figure 1b. Our regression models yielded an R^2 value of 0.51.

3.3 Gaps in the Recording due to Eye-Tracker

When visualizing the data, we found jumps in time that do not follow the 8 ms gap of the 125 Hz recording. Jumps in time between samples were on average 11.9 ms long with a standard deviation of 234 ms. In Figure 2, we visualize the number of seconds between two lines being logged into the different files in seconds that are over 8 ms (+ 10%). In total, the dataset contains 78.755 gaps in recording that are over this limit (m = 1324.9 ms, std = 4076.2 ms).

While sensors, such as eye trackers, deliver samples at a given frequency, the time is typically not precise. It can even happen that one or multiple consecutive samples are arbitrarily dropped. For this reason, we standardize the sampling frequency to perfect 125 Hz to counteract these gaps in the eye tracking data. Thus, we resampled the data to be exactly $1/frequency = 8msec$ apart. This allows us to investigate potential missing samples and gaps in the data and, thus, test infilling methods. During this process, we did not infill data; we merely ensured a perfect sample frequency of 125 Hz. If a sample was missing in the original dataset, the data was added at the correct time and marked as NaN for later processing using the different infilling methods.

4 RESULTS

In total, the dataset has 78.755 gaps in the recording where the eye tracker did not log any data and 52.219 sections of data where there were consecutive 0's (by us converted to NaN's) logged because the eye tracker was not able to recognize the pupil. We then resampled the data frame to allow for consistent steps in time according to the 125 Hz logging in the remaining of data. We then split the dataset in parts where there were more than 1500 consecutive NaN's in the dataset as the long short-term memory (LSTM) neural network (NN) provided by Annerer-Walcher et al. [2021] takes in a window of 1500 samples. In the initial dataset, having 1500 or more consecutive 0's happening only existed nine times; after our pre-processing, this happened 1198 times.

We infilled the dataset linearly and used cubic spline to fill in the remaining gaps in the data. For the cubic spline infilling, we take three samples at the start and end into consideration to allow the cubic spline interpolation to consider the velocity. If there was a gap of a single sample during the cubic spline interpolation, we linearly interpolated this, as cubic spline interpolation is not warranted in these scenarios. If there are NaN's in the 3 samples before or after the gap, we recusively go back or forward, respectively, until we find the allotted samples without NaN's. After interpolating, we had 3 datasets, one as provided and processed exactly by Annerer-Walcher et al. [2021], one linearly interpolated, and one cubic spline interpolated. These interpolation methods were applied over x and y screen-based coordinates for both eyes and the pupil dilation values for both eyes. Both of the interpolated datasets were without any missing samples or NaN's remaining.

In the work of Annerer-Walcher et al. [2021], they have used 157 participants from the whole dataset and excluded 9. Of this data, 135 sessions were used as training data to adjust the model parameters, and the remainder were used as testing data, pooling data from all tasks. Unfortunately, we could not identify in their publication or from the OSF page which participants were excluded and which sessions were used for which purposes. As such, we split the whole dataset, using a fixed seed, based on all participants (166), using 60% of the participants (99) for training, 20% for testing (33), and 20% for validation (33). We used the existing processing scripts after the infilling to generate the input for the LSTM model. We left the model's hyperparameters unchanged in the code.

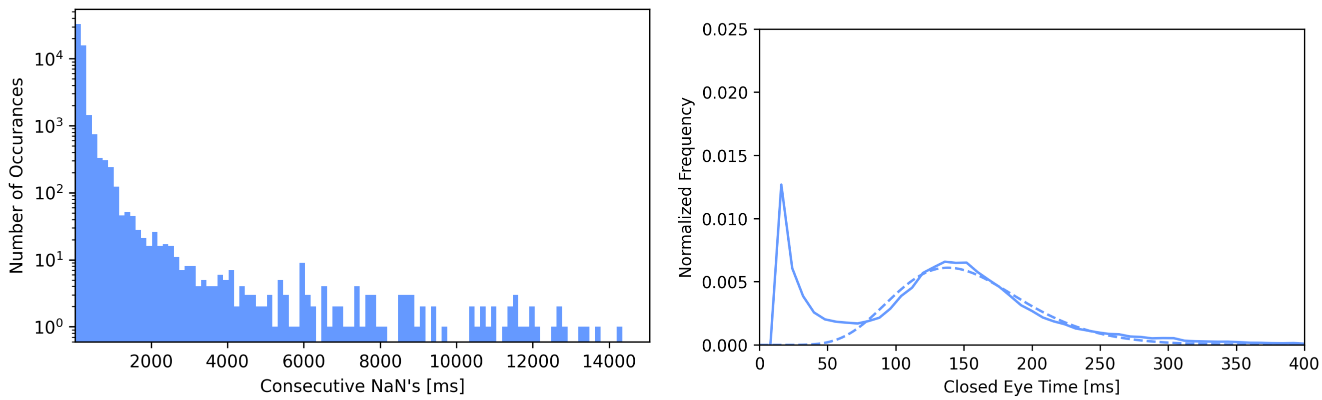
As such, our model starts with the input layer of 1500×16, followed by an LSTM layer of 64 units, after which there was a dense layer present of 64 units with a ReLu activation function with a dropout layer of 0.45 following the dense layer. After the dense layer, the model contains a convolutional layer followed by a max pooling layer as 1D. The final two layers contain another LSTM layer of 32 units and a final dense layer of 2 for the output. The model uses a nadam optimizer with a fixed learning rate of 0.001. The models were trained over 40 epochs in batches of 30 with an early stopping rule and a patience of 3 on the validation loss.

Our reference model yielded a training accuracy after five epochs of 0.8713, a test accuracy of 0.7699, and a validation accuracy of 0.7863. We achieved this with 6813 sets of windows for training, 2490 sets for testing, and 2176 sets for validation for 11,479 sets of windows. The model after linear interpolation yields a training accuracy after six epochs of 0.8834, a test accuracy of 0.7732, and a validation accuracy of 0.7946. We achieved this with 8765 sets of windows for training, 3157 sets for testing, and 2795 sets for validation for 14,717 sets of windows (28.2% more data over default). Our final model using spline interpolation yields a training accuracy after seven epochs of 0.9023, a test accuracy of 0.7555, and a validation accuracy of 0.7599. For this, we had 8732 sets of windows for training, 3141 for testing, and 2790 sets for validating (27.7% more data over default).

5 DISCUSSION

In this work, we applied two recommended interpolation methods from past work used to infill the missing data on a publicly available dataset. Our findings show that these interpolation methods over missing data points and gaps in data recording have major implications for the accuracy of the LSTM model provided with

[1] https://www.dpg.unipd.it/sites/dpg.unipd.it/files/BeGaze2.pdf, accessed 2024-04-05



(a) Consecutive NaN's **(b) Closed Eye Time**

Figure 1: (a) The log-scaled distribution of consecutive NaN's appearing the data. (b) We visualize the normalized frequency closed-eye time for the dataset. The dashed line represents an inverse Gaussian probability density function fitted to the data ($R^2 = 0.51$)

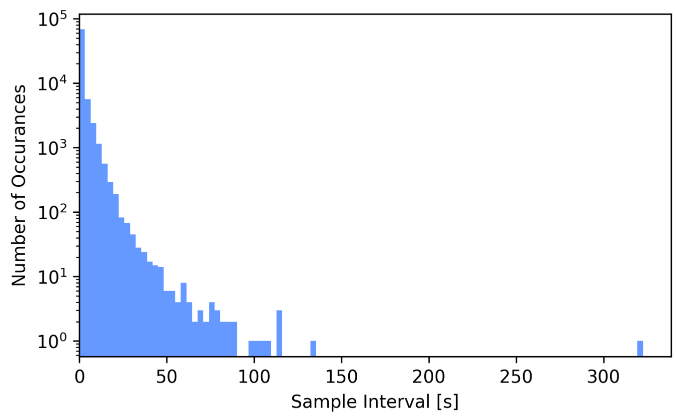
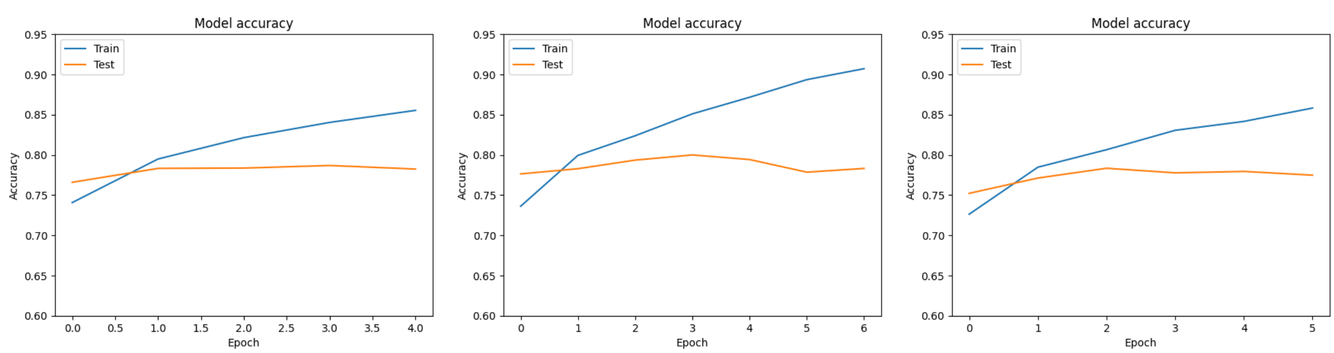
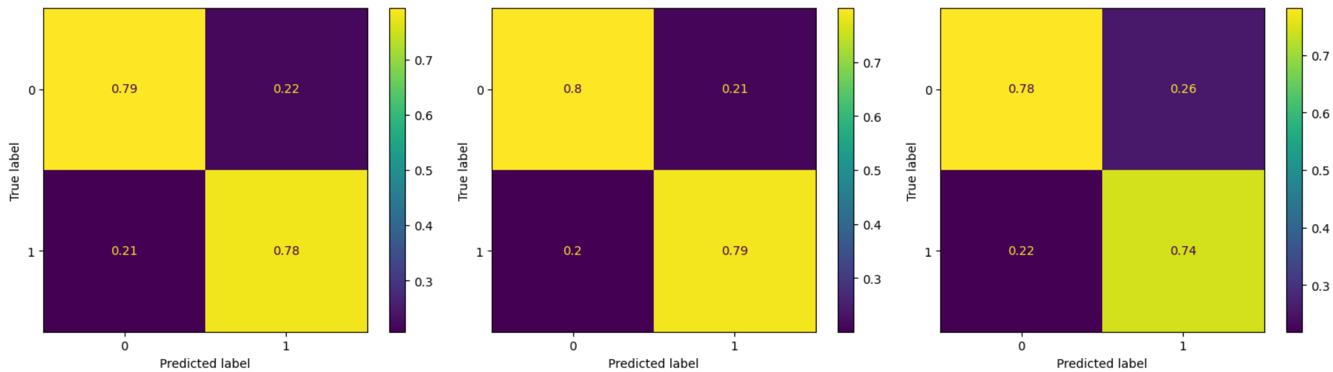


Figure 2: Illustration highlighting the interval (time between) of the logged samples over 8.8 ms (which should be the normal interval of a 125 Hz recording + allowance of 10% in variation).



(a) Default **(b) Linearly Infilled** **(c) Cubic Spline Infilled**

Figure 3: The train and test accuracy for the three models we trained. (a), shows the training and validation accuracy plotted for the default dataset. (b) and (c) show these for the linearly interpolated data and cubic spline interpolated data, respectively.



(a) Default **(b) Linearly Infilled** **(c) Cubic Spline Infilled**

Figure 4: Confusion Matrices from all the predictions on the validation sets (20% of the participants) using the provided model structure and input length. (a), shows the confusion matrix for the default dataset. (b) and (c) show these for the linearly interpolated data and cubic spline interpolated data, respectively.

the dataset. To the point that overfitting happens if we keep the original model structure.

Overfitting happens due to a modeling error when a function is too closely aligned to a limited set of data points. As a result, the model is useful only for reference to the initial data set and not to other data sets. In other words, it is not generalizable. We can see this in Figure 3, where the models all have an increase in training accuracy, while the test accuracy is almost stationary. We expect this to be in part related to not being able to reproduce the results of the previous work completely, as the exclusion of participants was not available to us, and neither was the split in data used for training and test purposes. Furthermore, the work documentation suggests that the experiment was recorded at 250 Hz, while the published data set suggests this was recorded or down-sampled to 125 Hz.

Infilling data comes at the "cost" of knowing the future sample. Only gaps can be interpolated using infilling methods. However, in interactive systems, this is not always the case; e.g., during a blink, a future sample for interpolation is not known. For this, we will need to experiment with extrapolation methods. Extrapolation in interactive systems is nothing novel and has even been done using neuronal networks for touch input [Henze et al. 2017]. However, a future real-time implementation of our approach will need to address the challenges of the "unknown" future.

6 CONCLUSION

In conclusion, interpolation methods are powerful for handling missing data and gaps in eye-tracking studies. We argue that the recommended interpolation methods should be preferred over leaving gaps in the recording or removing that data, as these can have huge implications on the available data as highlighted in Grootjen et al. [2024]. Both linear and cubic spline interpolation provide ways to improve the provided model's accuracy. Here, tweaks could boost the accuracy even further without overfitting the model on the training data, as there is an increase of over 25% in data available for the model. Future work should investigate the generalizability to other datasets.

REFERENCES

Larry A Abel, B Todd Troost, and Louis F Dell'Osso. 1983. The effects of age on normal saccadic characteristics and their variability. *Vision research* 23, 1 (1983), 33–37. https://doi.org/10.1016/0042-6989(83)90038-X

Alejandro Acien, Aythami Morales, Ruben Vera-Rodriguez, and Julian Fierrez. 2020. Smartphone Sensors for Modeling Human-Computer Interaction: General Outlook and Research Datasets for User Authentication. In *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, Madrid, Spain, 1273–1278. https://doi.org/10.1109/COMPSAC48688.2020.00-81

Abdul Rafey Aftab, Michael von der Beeck, and Michael Feld. 2020. You Have a Point There: Object Selection Inside an Automobile Using Gaze, Head Pose and Finger Pointing. In *Proceedings of the 2020 International Conference on Multimodal Interaction (ICMI '20)*. Association for Computing Machinery, New York, NY, USA, 595–603. https://doi.org/10.1145/3382507.3418836

Sonja Annerer-Walcher, Simon M Ceh, Felix Putze, Marvin Kampen, Christof Körner, and Mathias Benedek. 2021. How reliably do eye parameters indicate internal versus external attentional focus? *Cognitive Science* 45, 4 (2021), e12977. https://doi.org/10.1111/cogs.12977

Luis M. Bergasa, Jesús Nuevo, Miguel Angel Sotelo, Rafael Barea, and M. Elena Lopez. 2006. Real-Time System for Monitoring Driver Vigilance. *IEEE Transactions on Intelligent Transportation Systems* 7, 1 (March 2006), 63–77. https://doi.org/10.1109/TITS.2006.869598

WP Blount. 1927. Studies of the movements of the eyelids of animals: blinking. *Quarterly Journal of Experimental Physiology: Translation and Integration* 18, 2 (1927), 111–125. https://doi.org/10.1177/001872089403600209

Gianni Bremer, Niklas Stein, and Markus Lappe. 2023. Machine Learning Prediction of Locomotion Intention from Walking and Gaze Data. *International Journal of Semantic Computing* 17, 01 (March 2023), 119–142. https://doi.org/10.1142/S1793351X22490010

Jeffrey B Brookings, Glenn F Wilson, and Carolyne R Swain. 1996. Psychophysiological responses to changes in workload during simulated air traffic control. *Biological psychology* 42, 3 (1996), 361–377. https://doi.org/10.1016/0301-0511(95)05167-8

Han Collewijn, Johannes Van Der Steen, and Robert M. Steinman. 1985. Human eye movements associated with blinks and prolonged eyelid closure. *Journal of neurophysiology* 54, 1 (1985), 11–27. https://doi.org/10.1152/jn.1985.54.1.11

Michael Collins, Rhonda Seeto, Louella Campbell, and Murray Ross. 1989. Blinking and corneal sensitivity. *Acta ophthalmologica* 67, 5 (1989), 525–531. https://doi.org/10.1111/j.1755-3768.1989.tb04103.x

Lorenza S Colzato, Heleen A Slagter, Wery PM van den Wildenberg, and Bernhard Hommel. 2009. Closing one's eyes to reality: Evidence for a dopaminergic basis of psychoticism from spontaneous eye blink rates. *Personality and Individual Differences* 46, 3 (2009), 377–380. https://doi.org/10.1016/j.paid.2008.10.017

Michael J Doughty and Taher Naase. 2006. Further analysis of the human spontaneous eye blink rate by a cluster analysis-based approach to categorize individuals with 'normal'versus 'frequent'eye blink activity. *Eye & contact lens* 32, 6 (2006), 294–299. https://doi.org/10.1097/01.icl.0000224359.32709.4d

Heiko Drewes and Albrecht Schmidt. 2007. Interacting with the Computer Using Gaze Gestures. In *Human-Computer Interaction – INTERACT 2007*. Springer Berlin Heidelberg, Berlin, Heidelberg, 475–488.

Inger Ekman, Antti Poikola, Meeri Mäkäräinen, Tapio Takala, and Perttu Hämäläinen. 2008. Voluntary Pupil Size Change as Control in Eyes Only Interaction. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications* (Savannah, Georgia) *(ETRA '08)*. Association for Computing Machinery, New York, NY, USA, 115–118. https://doi.org/10.1145/1344471.1344501

Julie Epelboim and Patrick Suppes. 2001. A model of eye movements and visual working memory during problem solving in geometry. *Vision research* 41, 12 (2001), 1561–1574. https://doi.org/10.1016/S0042-6989(00)00256-X

Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015. Orbits: Gaze Interaction for Smart Watches using Smooth Pursuit Eye Movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 457–466. https://doi.org/10.1145/2807442.2807499

Irving Fatt and Barry A Weissman. 2013. *Physiology of the Eye: An Introduction to the Vegetative Functions.* Butterworth-Heinemann, Oxford, United Kingdom. https://books.google.de/books?id=H2OfAgAAQBAJ

Javier Galbally, Sebastien Marcel, and Julian Fierrez. 2014. Biometric Antispoofing Methods: A Survey in Face Recognition. *IEEE Access* 2 (2014), 1530–1552. https://doi.org/10.1109/ACCESS.2014.2381273

Jesse W. Grootjen, Henrike Weingärtner, and Sven Mayer. 2023. Highlighting the Challenges of Blinks in Eye Tracking for Interactive Systems. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications* (Tubingen, Germany) *(ETRA '23)*. Association for Computing Machinery, New York, NY, USA, Article 63, 7 pages. https://doi.org/10.1145/3588015.3589202

Jesse W. Grootjen, Henrike Weingärtner, and Sven Mayer. 2024. Uncovering and Addressing Blink-Related Challenges in Using Eye Tracking for Interactive Systems. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (2024-01-01) *(CHI Ź4)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3613904.3642086

Nishan Gunawardena, Michael Matscheko, Bernhard Anzengruber, Alois Ferscha, Martin Schobesberger, Andreas Shamiyeh, Bettina Klugsberger, and Peter Solleder. 2019. Assessing Surgeons' Skill Level in Laparoscopic Cholecystectomy Using Eye Metrics. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications (ETRA '19)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3314111.3319832

Ahmad Hassan, Wei Fan, Xiaoyu Hu, Wenhao Wang, and Hanxi Li. 2022. LSTM-based eye-movement trajectory analysis for reading behavior classification. In *International Conference on Image, Signal Processing, and Pattern Recognition (ISPP 2022)*, Michael Opoku Agyeman and Seppo Sirkemaa (Eds.), Vol. 12247. International Society for Optics and Photonics, SPIE, 1224715. https://doi.org/10.1117/12.2636952

Niels Henze, Sven Mayer, Huy Viet Le, and Valentin Schwind. 2017. Improving software-reduced touchscreen latency. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) *(MobileHCI '17)*. Association for Computing Machinery, New York, NY, USA, Article 107, 8 pages. https://doi.org/10.1145/3098279.3122150

Javier Hernandez-Ortega, Julian Fierrez, Aythami Morales, and Javier Galbally. 2019a. Introduction to Face Presentation Attack Detection. In *Handbook of Biometric Anti-Spoofing*. Springer International Publishing, Cham, 187–206. https://doi.org/10.1007/978-3-319-92627-8_9 Series Title: Advances in Computer Vision and Pattern Recognition.

Javier Hernandez-Ortega, Shigenori Nagae, Julian Fierrez, and Aythami Morales. 2019b. Quality-based Pulse Estimation from NIR Face Video with Application to Driver Monitoring. http://arxiv.org/abs/1905.06568 arXiv:1905.06568 [cs, eess].

K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. van de Weijer. 2011. *Eye Tracking: A comprehensive guide to methods and measures.* OUP Oxford, Oxford, United Kingdom. https://books.google.de/books?id=5rIDPV1EoLUC

Aristea Ladas, Christos Frantzidis, Panagiotis Bamidis, and Ana B. Vivas. 2014. Eye Blink Rate as a biological marker of Mild Cognitive Impairment. *International Journal of Psychophysiology* 93, 1 (July 2014), 12–16. https://doi.org/10.1016/j.ijpsycho.2013.07.010

Sharon Leal and Aldert Vrij. 2008. Blinking During and After Lying. *Journal of Nonverbal Behavior* 32, 4 (Dec. 2008), 187–194. https://doi.org/10.1007/s10919-008-0051-0

Lars Lischke, Valentin Schwind, Kai Friedrich, Albrecht Schmidt, and Niels Henze. 2016. MAGIC-Pointing on Large High-Resolution Displays. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 1706–1712. https://doi.org/10.1145/2851581.2892479

Jean-Luc Lugrin, Dennis Wiebusch, Marc Erich Latoschik, and Alexander Strehler. 2013. Usability benchmarks for motion tracking systems. In *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology* (Singapore) *(VRST '13)*. Association for Computing Machinery, New York, NY, USA, 49–58. https://doi.org/10.1145/2503713.2503730

I. Scott MacKenzie and Colin Ware. 1993. Lag as a determinant of human performance in interactive systems. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (Amsterdam, The Netherlands) *(CHI '93)*. Association for Computing Machinery, New York, NY, USA, 488–493. https://doi.org/10.1145/169059.169431

Samantha Mann, Aldert Vrij, and Ray Bull. 2002. Suspects, lies, and videotape: An analysis of authentic high-stake liars. *Law and Human Behavior* 26, 3 (June 2002), 365–376. https://doi.org/10.1023/A:1015332606792

Almudena Palacios-Ibáñez, Javier Marín-Morales, Manuel Contero, and Mariano Alcañiz. 2023. Predicting Decision-Making in Virtual Environments: An Eye Movement Analysis with Household Products. *Applied Sciences* 13, 12 (2023). https://doi.org/10.3390/app13127124

Gang Pan, Lin Sun, Zhaohui Wu, and Shihong Lao. 2007. Eyeblink-based Anti-Spoofing in Face Recognition from a Generic Webcamera. In *2007 IEEE 11th International Conference on Computer Vision*. IEEE, Rio de Janeiro, Brazil, 1–8. https://doi.org/10.1109/ICCV.2007.4409068

Sudi Patel, Ross Henderson, L Bradley, B Galloway, and L Hunter. 1991. Effect of visual display unit use on blink rate and tear stability. *Optom Vis Sci* 68, 11 (1991), 888–892. https://doi.org/10.1097/00006324-199111000-00010

L. Perreault, Bernard Bobée, and Peter Rasmussen. 1999. Halphen Distribution System. I: Mathematical and Statistical Properties. *Journal of Hydrologic Engineering - J HYDROL ENG* 4 (07 1999). https://doi.org/10.1061/(ASCE)1084-0699(1999)4:3(189)

Ken Pfeuffer and Hans Gellersen. 2016. Gaze and Touch Interaction on Tablets. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 301–311. https://doi.org/10.1145/2984511.2984514

Tuan D Pham. 2021. Time–frequency time–space LSTM for robust classification of physiological signals. *Scientific reports* 11, 1 (2021), 6936. https://doi.org/10.1038/s41598-021-86432-7

Niklas Stein, Gianni Bremer, and Markus Lappe. 2022. Eye Tracking-based LSTM for Locomotion Prediction in VR. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, New York, NY, USA, 493–503. https://doi.org/10.1109/VR51125.2022.00069

John A Stern, Donna Boyer, and David Schroeder. 1994. Blink rate: a possible measure of fatigue. *Human factors* 36, 2 (1994), 285–297.

Yi-Fang Tsai, Erik Viirre, Christopher Strychacz, Bradley Chase, and Tzyy-Ping Jung. 2007. Task performance and eye activity: predicting behavior relating to cognitive workload. *Aviation, space, and environmental medicine* 78, 5 (2007), B176–B185.

Jayson Turner, Andreas Bulling, Jason Alexander, and Hans Gellersen. 2014. Cross-device gaze-supported point-to-point content transfer. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Safety Harbor, Florida) *(ETRA '14)*. Association for Computing Machinery, New York, NY, USA, 19–26. https://doi.org/10.1145/2578153.2578155

Karl F Van Orden, Tzyy-Ping Jung, and Scott Makeig. 2000. Combined eye activity measures accurately estimate changes in sustained visual task performance. *Biological psychology* 52, 3 (2000), 221–240. https://doi.org/10.1016/S0301-0511(99)00043-5

Manhua Wang, Seul Chan Lee, Harsh Kamalesh Sanghavi, Megan Eskew, Bo Zhou, and Myounghoon Jeon. 2021. In-Vehicle Intelligent Agents in Fully Autonomous Driving: The Effects of Speech Style and Embodiment Together and Separately. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '21)*. Association for Computing Machinery, New York, NY, USA, 247–254. https://doi.org/10.1145/3409118.3475142

P Wolkoff, Jacob Klenø Nøjgaard, P Troiano, and B Piccoli. 2005. Eye complaints in the office environment: precorneal tear film integrity influenced by eye blinking efficiency. *Occupational and environmental medicine* 62, 1 (2005), 4–12. https://doi.org/10.1136/oem.2004.016030

Guanhua Zhang, Susanne Hindennach, Jan Leusmann, Felix Bühler, Benedict Steuerlein, Sven Mayer, Mihai Bâce, and Andreas Bulling. 2022. Predicting Next Actions and Latent Intents during Text Formatting. In *Proceedings of the CHI Workshop Computational Approaches for Understanding, Generating, and Adapting User Interfaces* (2022-01-01). Association for Computing Machinery, New York, NY, USA, 1–6. https://sven-mayer.com/wp-content/uploads/2022/08/zhang2022predicting.pdfhttps://perceptualui.org/publications/zhang22_caugaui/

Yanxia Zhang, Ken Pfeuffer, Ming Ki Chong, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2017. Look together: using gaze for assisting co-located collaborative search. *Personal and Ubiquitous Computing* 21, 1 (01 Feb 2017), 173–186. https://doi.org/10.1007/s00779-016-0969-x