Towards Pupil-Assisted Target Selection in Natural Settings: Introducing an On-Screen Keyboard

Christoph Strauch, Lukas Greiter, Anke Huckauf

HAL Id: hal-01717215
https://hal.inria.fr/hal-01717215

Submitted on 26 Feb 2018

Towards Pupil-Assisted Target Selection in Natural Settings: Introducing an On-screen Keyboard

Christoph Strauch, Lukas Greiter, Anke Huckauf

Department of General Psychology, Ulm University, Ulm, Germany

{christoph.strauch, lukas.greiter, anke.huckauf}@uni-ulm.de

Abstract. Preliminary reports have shown the possibility to assist input commands in HCI via pupil dilation. Applicability of these findings is however subject to further investigations, since the specificity of changes in diameter is low, e.g. through variations in brightness. Investigating employability and shape of pupil size dynamics outside a strictly controlled laboratory, we implemented the emulation of selection via an integrated mechanism of pupil dilation and constriction that could speed up a dwell time of 1.5 s. During the operation of an on-screen keyboard, 21 subjects were able to type via this mechanism, needing 1 s on average per keystroke and producing only slightly more than 1% false positive selections. Hereby, pupil dynamics were assessed. More than 90% of keystrokes could be accelerated under assistance of pupil variations. As suggested from basic research, pupil dilated when fixating later selected keys and constricted shortly afterwards. This finding was consistent between all subjects, however, pupil dynamics were shifted in regard to temporal occurrence and amplitude of diameter changes. Pupil-Assisted Target Selection shows potential in non-strictly controlled environments for computer input and may be further improved on the basis of this data. This might culminate in an integrated gaze-based object selection mechanism that could go beyond the benchmarking dwell time performance.

Keywords: eye typing; gaze-based interaction; physiological computing

1   Introduction

Brightness-independent pupil size changes are a variable that could assist object selection in computer input [1–3]. Hereby, especially implicit pupil size changes in response to selection, which might reveal user intention, are promising. Pupil size changes are associated with comparably short latencies, which enables high-resolution information retrieval. Moreover, changes require no cognitive effort of the user and are trackable remotely even with low-cost trackers [4]. A first interface evaluating the idea of employing pupil size changes as a mean to select in eyes only interaction, shows a promising performance, although a dwell time approach can't be outperformed. The authors report a specific dilation and a subsequent constriction that accompanies selections [1]. However, it is not yet clear whether the crucial signal components that have been iden-

tified in basic-research paradigms, can be reproduced if brightness is not strictly controlled, baselines thus differ, and when users interact repeatedly with interfaces in natural environments. With this paper, we are – to our knowledge – the first to present an emulated selection in a standard interface that was operated under assistance of pupil size changes, in this case implicit changes in size as a mean to operate an on-screen keyboard. We evaluate the keyboard next to a window and test whether pupil size dynamics, similar to these reported in basic-research, can be obtained during eye-typing.

2   Related Work

Regulated by two antagonistic muscle systems, the pupil determines the influx of light onto the retina analogously to the aperture of a camera onto the sensor. Similarly to the aperture, also the visual acuity is influenced by these changes. Variations in pupil size, however, are not only limited to visual adjustments of the eye, but can also indicate psychological processes. In comparison to other peripheral-physiological indicators, these changes occur fast and are sensitive to even small changes in activity. Pupil size changes indicate a broad variety of psychological processes, ranging from cognitive effort [5], to emotional activation [6], visual search [7], or attentional processes [8, 9]. This variety is boon and bane to researchers simultaneously: on the one hand, it enables the investigation of many factors in controlled experiments, on the other hand, the low specificity leads to a convoluted signal that makes it hard to conclude from. This low specificity hinders the application of pupil size changes in HCI up to now.

However, among these factors, some may lead to a pupillary response that could be specific, and also highly relevant to HCI. In addition to changes in arousal, e.g. elicited by emotional activation [6], pupil size can indicate activity of locus coeruleus and the norepinephrine system in almost perfect temporal synchrony. This nucleus is linked to, among others, attentional processes [10]. Einhäuser and colleagues [8] report that pupil size changes can indicate a selected number from a set of numbers and therefore may reveal decisions. De Gee et al. [11] describe larger pupil dilations, when subjects think that they have identified a signal for which they have to look out. Also, spotting the correct target in a visual search task is associated with increased pupil diameter [7, 12]. Fixations not on target cause no considerable increase in pupil diameter, while fixations on targets cause an increase of about 0.06 mm on average [13]. However, in another publication, Bednarik et al. [14] use a machine learning approach to investigate multiple object selection accompanying eye-based variables, including pupil size changes and state that they cannot replicate the endings of [13], still pupil size helps to predict object selection above chance [14]. A possible reason is seen in the applied character of the investigation that is in contrast to the basic research scenario reported by [13]. Unfortunately, no signal dynamics are given in [14]. When provided, signal courses are largely comparable among studies and show a fast dilation in response to the process that is taking place around the fixation of intended objects. Given the high similarity of signal courses, it seems that pupil size variations emerge as specific in paradigms that are comparable to the aforementioned and could therefore be identified easily. Moreover, constrictions that directly follow the dilation after selections are reported to carry

information, that is, bigger constrictions are associated with correct foregoing selections, compared to smaller constrictions associated with uncertain selections and erroneous selections [15].

As of today, three different pupil-based mechanisms are introduced that can reportedly be employed to select. Voluntarily, pupil size can be modulated by means of cognitive strategies, which leads to the idea to use pupil size expansion as an input mechanism to HCI [16]. Stoll et al. [17] demonstrate that voluntary changes in diameter could serve as a binary communication channel that enables severely disabled patients to say either yes or no via a computer. Ehlers et al. [3] build on this approach and show a live implementation of this idea, here, even a search and select task can be performed, although selection times are long. Another method builds on the phenomenon that pupil size oscillations can reveal covert attention to differentially flickering objects on a screen, hereby subjects can select between a limited number of letters [2]. Given the comparable long selection times and limited possibilities for application in standard interfaces, we focus on the implicit diameter changes described in the preceding paragraph. The identification of relevant objects and the decision for yes in comparison to no seems to lead to a pupil dilation. It could thus be assumed that the intention to select is reflected in pupil diameter. In a first interface building on this idea [1], users select from a set of nine circularly arranged objects in a visual search task. Objects are either selected via a dwell time, a specific increase in pupil size following a short dwell time, or either an increase in pupil size or a certain dwell time, whichever of both selects first. Similarly to the aforementioned studies on implicit changes, this investigation produced a pupil dilation of about 0.05 mm shortly after the later selected targets were fixated. It is concluded, that pupil size changes may assist gaze-based interaction and may be especially beneficial in combination with dwell times [1].

Eye-typing is a standard procedure to assess the eligibility of an input technique in eyes only interaction for evaluating the usability of computer operation [18]. However, there are application scenarios that are not covered by this method, such as the operation of public displays or mobile devices. A variety of methods have been proposed in order to increase the probability of predicting the user's intention correctly in eyes only interaction, in order to differentiate fixations with the intention to select from visual scanning to avoid the Midas-touch problem [19]. Those vary from eye-gestures [18], or smooth pursuits to moving objects [20], to dwell times, systematically prolonged fixation durations that are necessary to select [18, 19]. Key performance indicators for eye-typing comprise words typed per minute (wpm), with a word defined as five characters including space and punctuation, selection time, error rates, and keystrokes per character (KSPC). The best possible value for KSPC is 1, which results, if no character had to be corrected [21].

In order to investigate whether pupil size variations during object selections are comparable to these reported by e.g. [1, 13], whether they can be found even when selecting repeatedly in short time as in computer input, and to check for applicability of these changes in a natural environment, therefore, we implemented an on-screen keyboard that can be operated via a new combination of pupil size variations and dwell time. We are also the first to include the post-decisional constriction into an implicit pupil-based

selection and to report a pupil-assisted on-screen keyboard in a regular layout. Pupil size changes for every selected key are presented and the on-screen keyboard is evaluated next to a window in a user study.

Methods

2.1 System

Pupil size and gaze were tracked using a SMI iView X RED 120 eye-tracker (SensoMotoric Instruments GmbH, Teltow) mounted to the screen (DELL P2210 1650*1050 Pixels, 55 Hz). The eye tracker was automatically set to a sampling rate of 55 Hz. For both, implementation of the on-screen keyboard and data collection, PsychoPy version 1.84.02 was used [22]. Participants were seated 40 cm from the screen.

Fig. 1. Screenshot of the implemented on-screen keyboard. Gaze highlighted keys in light blue while a combination of pupil deviation criteria and fixation duration selected (highlighted by a dark blue frame).

The implemented on-screen keyboard is illustrated in Fig. 1. Keys except *space* and *backspace* had an on-screen size of 11.76 cm^2. The layout corresponded to the German QWERTZ-keyboard and contained keys for *dot*, *comma*, *space*, and *backspace* in addition to the letters of the alphabet, and thus contained 33 selectable keys. Written text and cursor were displayed centered above the keyboard. Line breaks shaped text legible also for longer input. Keys were highlighted in light blue, when entered by gaze. Successful selection was indicated by a dark blue frame around the key for 0.2 s and the specific command being applied to the text field. For this study a non-commercial keyboard was chosen since it allowed to have full access on every code component. Words

per minute, selection times, error rates, and keystrokes per character (KSPC) were assessed in conjunction with the system-usability (SUS) scale, a common unipolar usability scale, ranging from zero to 100, where 100 marks the best possible evaluation [23]. Moreover, a binary variable tracked whether keystrokes were accelerated via the pupil criteria.

Characters were selected on the basis of a score, which integrated dwell time and pupil size changes. Pupil deviation criteria were determined based on a preliminary study and on data from previous empirical results [1]. An exemplary pupil size dynamic of a keystroke during the preliminary study is depicted in Fig. 2. in conjunction with the dilation and the constriction criteria. The score was computed as follows: A dwell timer, starting at zero at fixation, increased by one each frame. If a pupil dilation over 0.04 mm in a moving window of 0.36 s (equals 20 frames) was detected, the score was once increased by 25. A detected subsequent pupil constriction of more than 0.07 mm in a second moving window of 0.36 s increased the score again once by 25. A character was selected when the score exceeded 82, which implies that all characters could be selected within a range of a maximum 1.5 s (only dwell) and a minimal selection time of 0.73 s (dwell in addition to both pupil criteria). 1.5 s is higher than regular dwell times, which go up to 1 s for novice users [18], thus, for this study this long dwell time serves as a hedge that prevents from being unable to select. After each selection and whenever looking away from the current key, the score was set to zero.
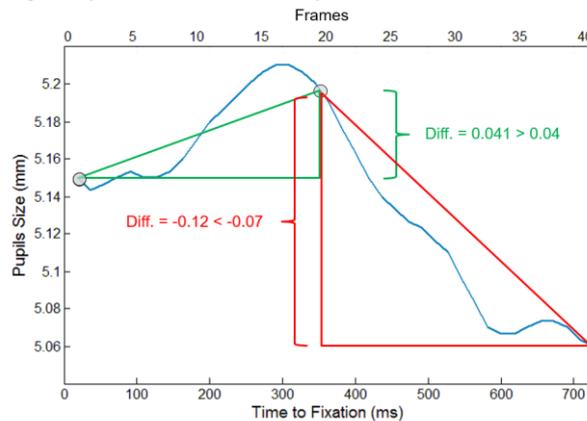

Fig. 2. Exemplary trial of an object selection, matching both, the dilation and the subsequent constriction criteria. Both scores plus 40 frames since fixation result in a score of 90, the key was thus selected at 727 ms.

2.2 Sample

A total of 21 users participated in the experiment, all of them were students at Anonymous University (female = 13, $M_{Age} = 22.76$). All participants reported normal or corrected to normal vision. All participants reported to not having consumed drugs prior to the experiment, neuronal diseases or traumatic brain injuries. Users signed an informed consent and took part in the study on a voluntary basis. They were partially

rewarded with course credits. Of 21 users, 15 were novices to gaze-based interaction. Six users took part in a previous study on gaze-based interaction but had no further experience.

2.3   Procedure

The study was carried out in a silent area next to a window. Brightness ranged from 45 lx to 3500 lx. After signing an informed consent, users were instructed about the on-screen keyboard and the task using a printed screenshot. If subjects remarked errors, they were asked to only correct those errors that are not more than two characters before the current letter, in order to ensure comparable numbers of characters and time typed among subjects. Then participants were asked to sit down in front of the eye tracker and where calibrated using the automatic nine-point calibration and validation by SMI. The task consisted of three sequential conditions. The pangram, a sentence containing every letter of the alphabet, *Franz jagt im komplett verwahrlosten Taxi quer durch Bayern* (Franz chases in the completely neglected taxi across Bavaria), the negatively connoted words *Selbstmord* (suicide) and *Albtraum* (nightmare), and the positively connoted words *Liebe* (love) and *Erfolg* (success) had to be typed successively using the on-screen keyboard. Emotional words were chosen to estimate whether these lead to differential signal dynamics, given the susceptibility of pupil diameter to emotionally elicited arousal e.g. [24]. Participants were not aware of the underlying selection criteria, but were instructed to select a focused character by merely intending to select. After writing all words, users were handed out a system-usability scale (SUS).

3   Results

On average, 91.6 letters, including space, were typed. After calibration, users needed $M = 241.46$ s on average to type the required text ($SD = 74.60$ s). $M = 4.55$ words were typed per minute ($SD = 1.24$). On average, users needed $M = 1.04$ s ($SD = 0.05$ s) from fixation of a key until its selection. False positive selections of keys accounted for M = 1.10 % of all key selections, clear spelling mistakes were not considered to be false positive. KSPC was $M = 1.02$ ($SD = 0.03$). User satisfaction, as assessed with the SUS, revealed an average satisfaction of $M = 77.64$ ($SD = 11.71$). Pupil size changes lowered the selection score in $M = 93.19$ % ($SD = 4.52$ %) of selections.

This finding suggests that pupil changes that fulfill at least the dilation criterion must have been present at more than 90% of all keys selected. However, this does not allow precise conclusions on the underlying signal. Signal dynamics can reveal, whether a common course can be found, a feature that is crucial for developing future selection criteria. Pupil dynamics accompanying all keystrokes were analyzed post-hoc with regard to the preceding and the following development. For the further analysis, the average of 700 ms to 600 ms prior to fixation of later selected keys was subtracted from every following data point as a local baseline. Signal shapes for emotional words and the neutral sentence showed no remarkable difference, all keystrokes were thus analyzed together. As illustrated in Fig. 2., every user showed a consistent double peaked

signal with the first local maximum in a window between 300 ms prior to key fixation until 700 ms post key fixation. Also, within participants the same signal dynamics can be found temporarily shifted. Signal dynamics thus vary in regard to both, amplitude and temporal occurrence. This could be explained by cognitive processes, such as the intention to select, which should only on average be linked to the onset of fixations.
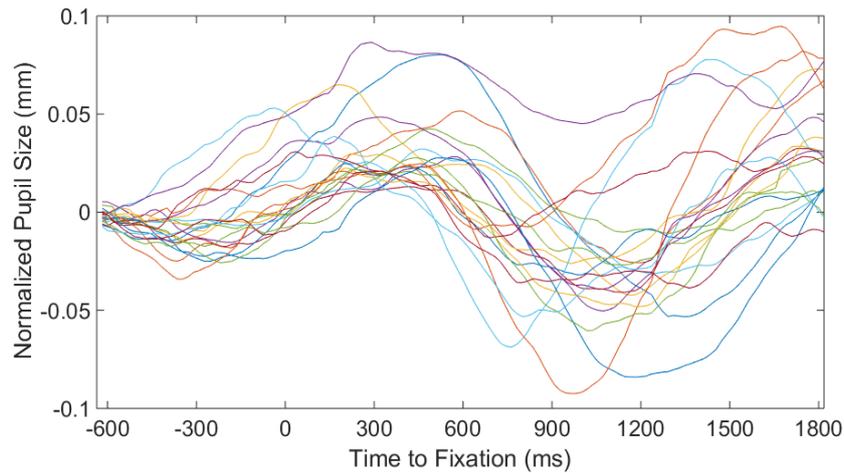

Fig. 2. Individually averaged user signal dynamics around the point of fixation of the later selected key (21 averages consisting of n = 1814 signal courses). Baseline corrected with a local, directly preceding baseline of 100 ms

Klingner [13] argues to reorient signal dynamics in order to be able to identify underlying signal patterns. This is crucial in order to be able to determine an estimate for the effect size of dilations and constrictions. All signal dynamics were consequently oriented at the local maximum between 300 ms prior to until 700 ms post fixation, a feature that could be observed in every subject (Fig. 3). Functional confidence intervals in accordance to [6] reveal that the averaged pupil size between subjects differs significantly from baseline mean until 500 ms post to the local maximum. Dilations go up 0.13 mm on average ($SD = 0.03$ mm).
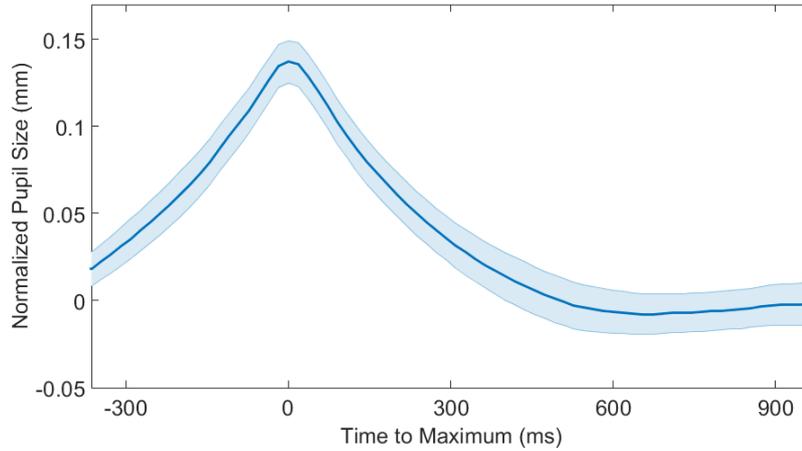
Fig. 3. Averaged signal dynamics of all participant means oriented at the local maximum of a window between 300 ms before until 700 ms after fixation. Only trials in which the maximum was not the first data point of the window were considered (21 averages, consisting of n = 1301 signal courses). Shaded bars mark functional 95% confidence intervals for every data point.

4   Discussion and Conclusion

Pupil size variations can be used as a mean to support dwell time based on-screen-keystrokes during computer input. We implemented a system that led to a user performance that is almost competitive to the performance of dwell time on-screen keyboards, operated by novices (employing slow dwell times of 1 s), in regard to selection times and error rates [18]. However, words per minute are still slightly below the average scores of novices writing with dwell time based interfaces [25]. KSPC were slightly better than for users typing with a novice dwell time of 876 ms [26]. User satisfaction, assessed with SUS can be described as usable [27]. However, for the previous findings, it has to be pointed out, that the comparability between studies on gaze-based interaction may be limited by between study factors such as accuracy of tracking or the sample of users [28]. Thus, in order to give a more stable estimation of the performance of this selection mechanism, a comparative dwell-time baseline condition is needed and should be addressed as a next step. For a comparison, a shorter dwell time adjusted to the time window of occurrences of the signal maxima might be applied. For over 90 % of keystrokes, pupil size changes lowered the dwell time. Thus, for now, pupil assisted eye-typing cannot outperform dwell time, but there is further potential for improvements. Keystrokes were highly consistently reflected in pupil size changes between every of the 21 users. Still, temporal phase and pupil dilation differed between subjects. Further analyses may reveal whether this variance effects system performance. The obtained signal dynamics correspond to signal dynamics reported in [1, 11, 13] and therefore replicate these findings outside the laboratory, in a much applied scenario, and with another eyetracker. When orienting the signals at the local maximum that is apparent

between 300 ms prior to and 700 ms post to fixation, it becomes evident that effect sizes are substantially larger than in the averaged key-fixation aligned data

The combination of pupil size variations and dwell as a mean to operate an on-screen keyboard was implemented on a self-developed keyboard. In a future investigation, this method could also be evaluated on existing commercial on-screen keyboards that have an optimal layout or on other interfaces, such as public displays to estimate the performance and usability compared to already existing input mechanisms. Since this approach may especially be beneficial to severely disabled users, it has to be evaluated to which extent these can use this setup, if neuronal diseases or brain injuries are present. It has to be stated that the integration of pupil diameter variation and fixation was chosen as a best estimation and is therefore subject to further improvements. The reported effect sizes may serve as an orientation for future pupil supported interfaces. Given the comparably large average effect size, it should be possible to find criteria that are robust to non-selection related variations in pupil diameter. Also, future interfaces should tailor criteria to the individual user, given the variance in maximum dilations between users. Bednarik et al. [14] only found a moderate prediction rate for pupil size changes during interaction. This might be explained by the controlled setting of the investigation or an effect that can only be found on average but not on single trial basis. Still, user performance and signal dynamics that were comparable for every participant seem promising. The shifted occurrence of the same signal pattern both between and within users, may be an indicator for cognitive processes, such as decisions, which differ in timing. However, from this experiment, signal dynamics in response to not intended objects cannot be derived. Orienting response [29] could partially explain the observed signal dynamics, still the keyboard was present permanently which should limit this effect. A small dilation in combination with a subsequent stronger constriction is also linked to fixations [30], however this finding might partially be explained with the orienting response [29] and/or visual search [7] and does not match the finding of [13].

This investigation collected data on pupil size variations during key selection. We are the first to describe an on-screen keyboard that could be operated using pupil size variations. Implementation and evaluation reveal a very high precision and selection times that are almost comparable to novice dwell times. Further room for adjustments might even allow to improve pupil assisted selection mechanisms beyond this level. Signal dynamics correspond to those obtained in controlled laboratory studies and were clearly found for every user, effect sizes in dilation are promising when aligning these dynamics.

5   References

1. C. Strauch, J. Ehlers, and A. Huckauf, "Pupil-Assisted Target Selection (PATS)," in *Proceedings of the 16th IFIP TC 13 International Conference on Human-Computer Interaction, INTERACT* . Springer International Publishing, 2017.
2. S. Mathôt, J.-B. Melmi, L. van der Linden, and S. van der Stigchel, "The mind-writing pupil: A human-computer interface based on decoding of covert attention through pupillometry," *PloS one*, vol. 11, no. 2, e0148805, 2016.
3. J. Ehlers, C. Strauch, and A. Huckauf, "A View to a Click: Pupil Size Changes as Input Command in Eyes-only Human-Computer Interaction," *International Journal of Human-Computer Studies*, submitted.
4. E. Dalmaijer, "Is the low-cost EyeTribe eye tracker any good for research?," *PeerJ PrePrints*, Jan. 2014.
5. E. H. Hess and J. M. Polt, "Pupil size in relation to mental activity during simple problem-solving," *Science*, vol. 143, no. 3611, pp. 1190–1192, 1964.
6. J. Ehlers, C. Strauch, J. Georgi, and A. Huckauf, "Pupil Size Changes as an Active Information Channel for Biofeedback Applications," *Applied Psychophysiology and Biofeedback*, no. 42, pp. 1–9, 2016.
7. C. M. Privitera, L. W. Renninger, T. Carney, S. Klein, and M. Aguilar, "Pupil dilation during visual target detection," *Journal of Vision*, vol. 10, no. 10, p. 3, 2010.
8. W. Einhäuser, C. Koch, and O. Carter, "Pupil dilation betrays the timing of decisions," *Frontiers in Human Neuroscience*, vol. 4, p. 18, 2010.
9. S. M. Wierda, H. van Rijn, N. A. Taatgen, and S. Martens, "Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution," *Proceedings of the National Academy of Sciences*, vol. 109, no. 22, pp. 8456–8460, 2012.
10. M. S. Gilzenrat, S. Nieuwenhuis, M. Jepma, and J. D. Cohen, "Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 10, no. 2, pp. 252–269, 2010.
11. de Gee, Jan Willem, T. Knapen, and T. H. Donner, "Decision-related pupil dilation reflects upcoming choice and individual bias," *Proceedings of the National Academy of Sciences*, vol. 111, no. 5, E618-E625, 2014.
12. Low, T., Bubalo, N., Gossen, T., Kotzyba, M., Brechmann, A., Huckauf, A., & Nürnberger, A., "Towards Identifying User Intentions in Exploratory Search using Gaze and Pupil Tracking," in *Proceedings of the 2017 Conference on Human Information Interaction and Retrieval*: ACM, 2017.
13. J. Klingner, "Fixation-aligned pupillary response averaging," in *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*: ACM, 2010, pp. 275–282.
14. R. Bednarik, H. Vrzakova, and M. Hradis, "What do you want to do next: a novel approach for intent prediction in gaze-based interaction," in *Proceedings of the Symposium on Eye Tracking Research and Applications*: ACM, 2012, pp. 83–90.
15. A. E. Urai, A. Braun, and T. H. Donner, "Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias," *Nature Communications*, vol. 8, 2017.
16. I. Ekman, A. W. Poikola, and M. K. Mäkäräinen., "Invisible eni: Using gaze and pupil size to control a game," in *Proceedings of the 2008 CHI Conference on Human Factors in Computing Systems*: ACM, 2008.
17. J. Stoll *et al.,* "Pupil responses allow communication in locked-in syndrome patients," *Current Biology*, vol. 23, no. 15, R647-R648, 2013.
18. P. Majaranta, *Text entry by eye gaze*: Tampereen yliopisto, 2009.

19. R. J. K. Jacob, "The use of eye movements in human-computer interaction techniques: what you look at is what you get," *ACM Transactions on Information Systems (TOIS)*, vol. 9, no. 2, pp. 152–169, 1991.

20. M. Vidal, A. Bulling, and H. Gellersen, "Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets.," in *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*: ACM, 2013.

21. R. W. Soukoreff and I. S. MacKenzie, "Metrics for text entry research: An evaluation of MSD and KSPC, and a new unified error metric," in *Proceedings of the 2003 CHI Conference on Human Factors in Computing Systems*, 2003.

22. J. W. Peirce, "PsychoPy—psychophysics software in Python," *Journal of Neuroscience methods*, vol. 162, no. 1, pp. 8–13, 2007.

23. J. Brooke, "SUS-A quick and dirty usability scale," *Usability Evaluation in Industry*, vol. 189, no. 194, pp. 4–7, 1996.

24. T. Partala and V. Surakka, "Pupil size variation as an indication of affective processing," *International Journal of Human-Computer Studies*, vol. 59, no. 1, pp. 185–198, 2003.

25. P. Majaranta and K.-J. Räihä, "Text entry by gaze: Utilizing eye-tracking," *Text entry systems: Mobility, accessibility, universality*, pp. 175–187, 2007.

26. P. Majaranta, U. K. Ahola, and O. Spakov, "Fast eye-typing with an adjustable dwell time," in *Proceedings of the 2009 CHI Conference on Human Factors in Computing Systems*: ACM, 2009.

27. J. R. Lewis and J. Sauro, "The factor structure of the system usability scale," in *International Conference on Human Centered Design*, 2009.

28. K.J. Räihä, "Life in the Fast Lane: Effect of Language and Calibration Accuracy on the Speed of Text Entry by Gaze." in *Proceedings of the 15th IFIP TC 13 International Conference on Human-Computer Interaction, INTERACT* . Springer International Publishing, 2015.

29. C. Wang and D. P. Munoz, "Modulation of stimulus contrast on the human pupil orienting response," *European Journal of Neuroscience*, vol. 40, no. 5, pp. 2822–2832, 2014.

30. S. Mathôt, J.-B. Melmi, and E. Castet, "Intrasaccadic perception triggers pupillary constriction," *PeerJ*, vol. 3, e1150, 2015.