Playing Jazz with the Pupil Accommodative Response: A Novel Unexplored Pupil-Based Interaction Mode

Publisher:

Springer Nature

Published version:

DOI:10.1007/978-3-031-83845-3_11

Playing jazz with the Pupil Accommodative Response

First Author[1], Second Author[2], and Third Author[3]

[1] Institution 1
[2] Institution 2
[3] Institution 3

Abstract. In recent years, the Pupil Accommodative Response (PAR) has emerged as a promising communication strategy in Human Computer Interfaces (HCI) and augmentative and alternative communication devices for several reasons, including it being a repetitive, high-magnitude and low-noise innate response. Previous studies exploited the far-to-near focus switch that induces the PAR to extract a binary output. In the preliminary study that we are now presenting, we have opened up the possibility of detecting intermediate levels of response, with the aim of extracting a non-binary output from pupil size variations induced by shifts in focus between multiple targets. In the current context, we apply this strategy to a music machine, where the pupil size is continuously monitored and converted into musical notes resulting in a jazz melody. In this article we aim at presenting the preliminary results we obtained with our system and at discussing the challenges and limitations of this type of pupil-based application, that can be employed not only for entertainment purposes, but also to increase the user's awareness with respect to the physiological function and the voluntary control of pupil size. In perspective, the approach adopted for the music machine may be exploited in pupil-based HCIs to achieve higher information transfer rates.

Keywords: Music and entertainment · Pupil Accommodative Response · Human Computer Interfaces.

1 Introduction

In the context of Human Computer Interfaces (HCIs) and of Augmentative and Alternative Communication (AAC) devices for patients in Locked-In-State (LIS) and Complete-Locked-In-State (CLIS), the usage of voluntary control of pupillary size has recently emerged. This channel of communication is particularly interesting as pupil size is controlled by the autonomic nervous system, through smooth muscles, that seem to be relatively preserved in degenerative diseases, preferentially affecting motoneurons, like amyotrophic lateral sclerosis (ALS) [1, 2].

Several strategies can be exploited to generate variations of pupillary diameter; for instance stress [3], mental calculations [4, 5], feature-based attention[6], light stimuli [7] and emotions [8, 9]. In recent years, the possibility of voluntary driving pupil size by exploiting the Pupil Accommodative Response (PAR) has been explored with fair success [10–13]. The PAR consists in the constriction of the pupil, which occurs alongside the increase in lens curvature and convergence movements of the eyes whenever we shift focus from a distant to a near visual target, and vice-versa. This combination of three actions is referred to as the accommodation triad [14]. Consistent pupil constrictions can thus be achieved by voluntarily shifting the focus from a far to a near target, and no eye movement is required as far as the targets are aligned with the gaze (convergence movements occur in the other eye only). As compared to other signals commonly used in HCIs, e.g., EEG signals, the PAR presents a number of relevant advantages, presenting low-noise and a large and repeatable magnitude, thus allowing for easy detection. Also, importantly, it does not require particular training as it is an innate response.

For patients in LIS, who can effectively use eye-tracking based AACs, this communication channel is relatively slow and little appealing. Conversely, in the CLIS condition patients might have undergone a serious cognitive decline that may impair attempts to re-establish a new communication pathway with the outside world. As a consequence, it seems crucial to motivate and engage LIS patients with new HCIs before they may fall into CLIS, which calls for developing more appealing and effective AAC devices and applications.

With reference to pupil based communication it should be observed that the HCI user cannot see its own pupil, and it is thus important to gain awareness of its phyhsiological behaviour, particularly in terms of latency of response with respect to the intention. In this context, the possibility of increasing the engagement of the user with entertaining and appealing applications may also serve this purpose.

In the past, we developed a Graphical User Interface (GUI) that allowed the user to answer “yes” or “no” by moving the focus from a further to a closer target [10] and a pupil-controlled gaming application that allowed the user to drive a spaceship against moving satellites [11]. We now propose a music machine that offers a greater range of possibilities of configuration and usage.

In this article we aim at presenting some preliminary results and discussing limits and difficulties that this type of pupil-based application may face, along with intriguing possible variants.

2 Music machine’s block diagram

The music machine includes several key components including: a video camera, a video acquisition and processing board which detects and monitors the pupil size, an algorithm that converts changes in pupil size to changes in pitch of a sound and a loud speaker that presents the progressively generated melody. In addition, the series of visual targets is necessary to help the subject adjust the visual focus to different depths, as depicted in Fig1. In continuity with our previous studies [10–13] we decided to use the pupil's area as a measure of its size.

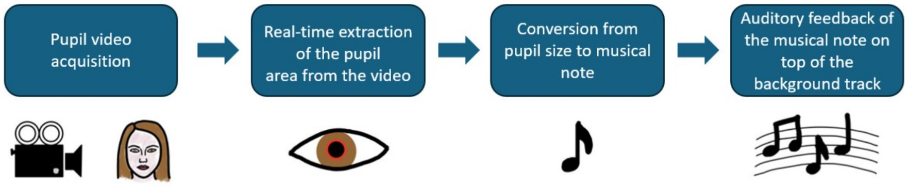

Fig. 1. Block diagram of the music machine. First pupil videos are acquired, then processed to extract the pupil size. This measure is converted into a musical note which is played by the system.

For the current implementation of the device we maintained a low-cost approach, as adopted in previous pupil-based applications [10–13]. Pupil size data are extracted from video data acquired with a 2.0 MP 1080p USB IR camera (Shenzen DingDang Smart Technology Co., Ltd., Shenzen, China) positioned on a desk at about 25 cm from the face of the user. Video data are processed firstly with MediaPipe[15], to automatically detect the circle enclosing the iris and cropping it as region of interest (ROI) from the rest of the frame, and then with image processing techniques to isolate the pupil from the frame and measure its pixel size, namely image equalization, binarization, thresholding, connected component analysis and ellipse fitting. To facilitate pupil detection, an IR led is employed for contrast enhancement. The GUI was developed in Python and was run on a single board computer (SBC) Raspberry Pi 4 model B, 4 GB RAM (Raspberry Pi Foundation, Cambridge, UK) Raspberry Pi5 to enhance system mobility. Prior to starting the music machine, a calibration phase is needed to measure the highest and the lowest pupil area by asking the user to alternatively focus on the furthest and on the closest target respectively (among a series of visual targets located in front of the user). The pupil size range is thus obtained for that subject in those environmental conditions. This range is then divided in to n equal intervals and each interval is assigned to a different note in progressive order, e.g., increasing note pitch assigned to decreasing pupil size. In this way all users will play notes in the same pitch range. Alternatively, notes could be assigned to fixed intervals within and over the full 2-8 mm physiological range [16]. In this way user would play over different pitch ranges depending on, e.g., age (older people tend to have smaller pupils [17]) and ambient light (smaller pupils in bright environments).

Pupil size is continuously monitored and converted into a musical note according to its current size, that the system makes audible by playing a previously recorded note from a musical instrument, in our case a saxophone. Notes are played over a jazz track that runs in background, as if the user was improvising its own musical solo. Furthermore, the set of available notes belongs to a pentatonic scale, which makes the resulting melody fit easily with the background track, also because the duration of the played notes matches its rhythm. Through these strategies we aim at making the music machine more entertaining and pleasant to listen to.

A series of visual targets is available in front of the user as a cue to help shifting the focus at different depths. Targets should be transparent or partially aligned with the subject's gaze of the dominant eye, so that movements of that eye are minimized avoiding disturbance of the pupil size assessment. Graphical details on the visual target may facilitate capturing the user's attention and improve the focusing task. Optimal distance between targets is still being investigated. In the last setting, 5 targets were spaced following a pseudo-logarithm scale, from 10 cm to 160 cm. This choice was made based on preliminary results and data from the literature showing that variations in pupil size changes are more prominent in the proximal than in the distal range [17–19]. We thus modeled the relationship between pupil size and target distance with a concave-down curve (qualitatively represented in Fig2). The association between pupil size and note made by dividing the pupil range into equal intervals has been arbitrarily chosen on the basis of preliminary experiments. With this choice, the user needs to keep in mind that notes are not uniformly distributed in space but are instead concentrated in the more proximal range, as depicted in Fig3 and discussed in Section 3. Note that, in the present implementation, targets are only meant as visual cues to help the subject to adjust the focus at variable depths, as represented in Fig3, with no bi-univocal association between target and musical note.

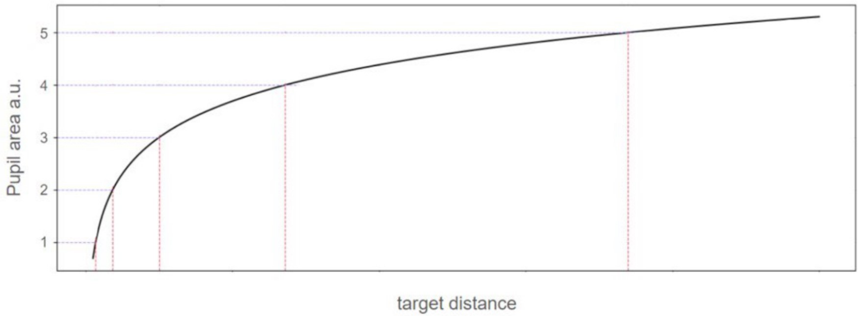

Fig. 2. Qualitative relationship between pupil size and target distance. Based on this curve, visual targets were positioned preferentially in the proximal range (see Fig3)

3 Discussion

During our preliminary experiments, several difficulties were encountered and several points are still to be addressed and investigated.

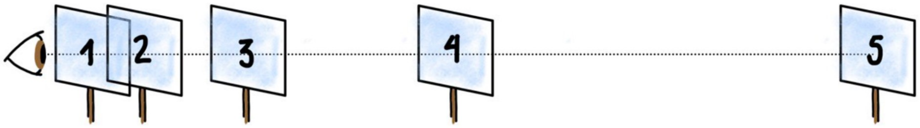

Fig. 3. Arrangement of visual targets. Semi-transparent targets, with visible details that facilitate gaze focus, are arranged with increasing inter-target distance, doubling at each step from left to right.

3.1 Confounding factors

First of all, eye blinking and light reflections introduced artifacts in the pupil size measurement. As eye blinking artifacts generated sudden and very fast drops and increases in pupil size, they were removed through a custom-made filtering algorithm [12]. Issues in light reflections were minimized by allowing for a manual selection of the binarization threshold, but they remain a limitation of the system as they might influence the performance of the image segmentation pipeline. Other confounding factors may include environmental lighting, stress and eye pathologies. Environmental lighting plays a role as it has been shown that the pupil accommodative response is influenced by the starting diameter of the pupil [17]. However, while the PAR is essentially preserved in moderately bright indoor environments [12], sudden changes in illumination may cause unwanted changes in pupil size. In a similar way, stress can induce pupil dilation, hence the user should be in a calm and relaxed mood not to randomly affect the pupil size.

3.2 Repeatibility of the pupil response

It has been previously reported that repeatable and clearly detectable PARs are obtained when switching from a far to a near target [12]. However, the repeatability of the response when focusing at intermediate distances, has been little investigated and appears to present large intra- and inter individual variability and thus poor reliability [18]. In the present application this problem may be counteracted by further increasing the width of pupil size intervals on the right side (at large pupil size).

3.3 Slowness of the pupillary response

In pupil-based applications, it should be kept in mind that there is a delay between the intention to switch focus on the near target and the actual pupil response; there is first of all a latency in the response in the order of 300ms in the case of a visually stimulated PAR [18] and of 600ms in case of acoustic stimulation [12], secondly, there is a limited velocity of constriction/dilatation. In particular, a well known characteristic of the PAR is that pupil dilatation is slower than pupil constriction. Semmelow and Stark (1973) [20] reported the maximum speed to be 3 times slower in dilatation than in constriction. An example of this phenomenon is shown in a typical PAR recorded from a healthy subject and reported in Fig4. The user needs to get acquainted with this feature as it limits their performance with the music machine, imposing a limit to the velocity of attaining the lowest pitch notes (associated with the large pupil size), after high-pitch ones.

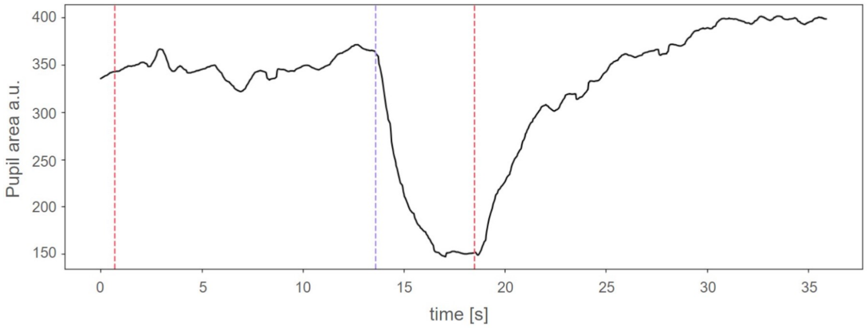

Fig. 4. Example of a pupil accommodative response in one subject initially focusing on a far target at 160 cm from the eye (starting from the red dashed line) to a near (10 cm) target (violet line) for 5 s, and back to the far target (red line). The changes in focus were not externally triggered but were spontaneously initiated by the subject, who simultaneously pressed a button for setting the markers. Note that the constrictory phase is faster than the dilatory.

3.4 Extension of the visual depth range

Another point to be debated is the optimal distance of the targets from the user. At first, we tried to exploit long distances, up to 320cm, and placed the targets at equal distances from 20cm to 320cm from the user's face. We observed that the pupil sizes attained when the user was asked to focus on the 320cm target or on the 245cm target were difficult to distinguish even when the subject was focusing on each target for a relatively long period of time (up to 30s). Conversely, pupil sizes attained on closer targets appeared more distinct: the more the user would look closely and the more the PARs are detectable and repeatable. This led to the choice of placing the further target not beyond 160cm, and concentrating the other targets in the proximal space. Lastly, we chose to make the distance between targets according to a pseudo-logarithmic progression, from 10 to 160cm from the user's face.

3.5 Fixed or minimal duration of the note

In our preliminary trials, we extracted the pupil areas for each frame collected by the IR camera, and updated the value of the note every 0.5s. Hence, notes had a fixed duration of 0.5s. We abandoned the idea of updating the note at each single frame because that would have resulted in possibly boring melodies, systematically made of sequences of adjacent notes. Alternatively, a minimal duration of the note, e.g., 3-400 ms could be introduced in order to preserve the possibility of playing non-adjacent notes (if pupil size changes are fast enough to pass through the adjacent interval within this time), while also granting the possibility of making notes of variable rather than fixed duration.

3.6 Discrete vs continuous visual cue

In the current implementation of the music machine we placed a number of visual cues at different distances (Fig3) to facilitate the focus switch of the user at different depths. However, alternative strategies can be adopted in terms of alignment, transparency and number of targets. For example, non-trasparent targets can be used if minimal gaze adjustments to switch among them can be accepted. A "continuous" target may also be envisaged, such as a single target expanding for the whole available depth (e.g. rigid or telescopic bar) that can be "followed" with minimal adjustments of the gaze. Possible differences between the two approaches still have to be investigated. In respective of the approach, we have to remember the relation between pupil size and target distance (Fig2). Consequently, the user should think of playing a "stretched" musical keyboard in which most of the available notes can be selected in the most proximal range (Fig5).

Fig. 5. According to the pupil size vs. target distance relationship presented above, and division of the pupil size range in equal intervals for note selection, the user must be aware that most available notes will be played when focusing in the proximal range. This has been visually represented with a piano whose keys increase in width, at increasing depth of focus

3.7 Active or passive playing

An intriguing variant of the music machine is presented in Fig6. In this case the user passively looks at a single visual target that held by a second (active) person. This latter may move the target more close to the user or more far and thus producing high-pitch or low-pitch notes according to the changes induced in the user's pupil. In other words the user has now become a passive instrument in the hands of the active player. This configuration would allow to present both discrete and continuous visual cues, depending on whether the target is always visible to the principal user or removed and replaced on the principal user's field of view.

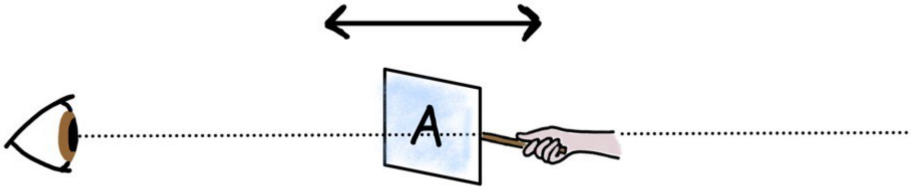

Fig. 6. "Passive" playing. In this case, the principal user would maintain the focus on the target while a second user moves it closer or further away from the subject.

4 Conclusion

The best configuration of this application will be attained once the relationship between pupil size and target distance will be better investigated. In particular, the repeatability of pupil size changes when focusing at intermediate depths needs to be characterized along with the relevance of target characteristics and environmental conditions. This musical application was devised for entertaining purposes for healthy people and patients but can be effectively employed also to increase the awareness about the physiological behaviour of the pupil. Lastly, a pivotal outcome of the current study would be opening up the possibility of extracting a non-binary output from the pupil accommodative response. In fact, in our previous studies [10–13] we only exploited the PAR in binary communication paradigms like providing YES/NO answers. The possibility to discriminate PAR at different depths of focus would allow to increase the information transfer rate in HCI applications and enhance the potential of this channel of communication.

References

1. Rojas, P., Ramírez, A. I., Fernández-Albarral, J. A., López-Cuenca, I., Salobrar-García, E., Cadena, M., and Ramírez, J. M. (2020). Amyotrophic lateral sclerosis: a neurodegenerative motor neuron disease with ocular involvement. Frontiers in Neuroscience, **14**, 566858 (2020).
2. Cozza, F., Lizio, A., Greco, L. C., Bona, S., Donvito, G., Carraro, E., and Lunetta, C. Ocular involvement occurs frequently at all stages of amyotrophic lateral sclerosis: Preliminary experience in a large Italian cohort. Journal of Clinical Neurology (Seoul, Korea), **17**(1), 96 (2021).
3. Graff, T. C., Luke, S. G., and Birmingham, W. C. Supportive hand-holding attenuates pupillary responses to stress in adult couples. PLoS One, **14**(2), e0212703 (2019).

4. Stoll, J., Chatelle, C., Carter, O., Koch, C., Laureys, S., and Einhäuser, W. Pupil responses allow communication in locked-in syndrome patients. Current biology, **23**(15), R647-R648 (2013).
5. Schuetz, I., Trojanek, J. , and Einhaeuser, W. Toward robust binary communication using pupil dilation by mental arithmetic. In EMICS '21: ACM CHI '21 Workshop on Eye Movements as an Interface to Cognitive State (2021).
6. Binda, P., Pereverzeva, M., and Murray, S. O. Pupil size reflects the focus of feature-based attention. Journal of neurophysiology, **112**(12), 3046-3052 (2014).
7. Mathôt, S., Melmi, J. B., Van Der Linden, L., and Van der Stigchel, S. The mind-writing pupil: a human-computer interface based on decoding of covert attention through pupillometry. PloS one, **11**(2), e0148805 (2016).
8. Partala, T., Jokiniemi, M., Surakka, V. Pupillary responses to emotionally provocative stimuli. In: Proceedings of the 2000 symposium on Eye tracking research and applications, pp. 123-129 (2000).
9. Weiss, M. W., Trehub, S. E., Schellenberg, E. G., and Habashi, P. Pupils dilate for vocal or familiar music. Journal of Experimental Psychology: Human Perception and Performance, **42**(8), 1061 (2016).
10. Chiarion, G., Ponzio, F., Terna, S., Moglia, C., Patti, E., Di Cataldo, S., Roatta, S. e-Pupil: IoT-based Augmentative and Alternative Communication device exploiting the pupillary near-reflex. IEEE Access, **10**, 130078-130088 (2022).
11. Cardinali, L., Roatta, S., Pertusio, R., Testa, M., Moglia, C.: Gaming for Training Voluntary Control of Pupil Size. Electronics, **11**(22), 3713 (2022).
12. Ponzio, F., Villalobos, L., Mesin, L., De'Sperati, C., and Roatta, S. A human-computer interface based on the "voluntary" pupil accommodative response. International Journal of Human-Computer Studies, **126**, 53-63 (2019).
13. Villalobos, L., Giusiano, S., Musso, L., de'Sperati, C., Riberi, A., Spalek, P., and Roatta, S. When assistive eye tracking fails: Communicating with a brainstem-stroke patient through the pupil accommodative response–A case study. Biomedical Signal Processing and Control, **67**(102515) (2021).
14. McDougal, D. H., and Gamlin, P. D. Autonomic control of the eye. Comprehensive physiology, **5**(1), 439 (2015).
15. Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., et al. Mediapipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172 (2019).
16. Spector RH. Clinical Methods: The History, Physical, and Laboratory Examinations, Chapter 58: The Pupils. 3rd edition. Walker HK, Hall WD, Hurst JW editors, Boston (1990).
17. Schaeffel, F., Wilhelm, H., and Zrenner, E. Inter-individual variability in the dynamics of natural accommodation in humans: relation to age and refractive errors. The Journal of Physiology, **461**(1), 301-320 (1993).
18. Kasthurirangan, S., and Glasser, A. Characteristics of pupil responses during far-to-near and near-to-far accommodation. Ophthalmic and Physiological Optics, **25**(4), 328-339 (2005).
19. Charman, W. N., and Radhakrishnan, H. Accommodation, pupil diameter and myopia. Ophthalmic and Physiological Optics, **29**(1), 72-79 (2009).
20. Semmlow, J., and Stark, L. Pupil movements to light and accommodative stimulation: a comparative study. Vision Research **13**(6), 1087-1100 (1973).