e-Pupil: IoT-based Augmentative and Alternative Communication device exploiting the pupillary near-reflex

IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 17 October 2022, accepted 11 December 2022, date of publication 14 December 2022, date of current version 19 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3229143

RESEARCH ARTICLE

e-Pupil: IoT-Based Augmentative and Alternative Communication Device Exploiting the Pupillary Near-Reflex

GIOVANNI CHIARION[1], FRANCESCO PONZIO[2], (Member, IEEE), STEFANO TERNA[3], CRISTINA MOGLIA[4], EDOARDO PATTI[5], (Member, IEEE), SANTA DI CATALDO[5], (Member, IEEE), AND SILVESTRO ROATTA[4]

[1]Department of Electronics and Telecommunications, Polytechnic University of Turin, 10129 Torino, Italy
[2]Interuniversity Department of Regional and Urban Studies and Planning, Polytechnic University of Turin, 10129 Torino, Italy
[3]TomorrowData, 10145 Turin, Italy
[4]Department of Neurosciences, "Rita Levi Montalcini", University of Turin, 10124 Torino, Italy
[5]Department of Control and Computer Engineering, Polytechnic University of Turin, 10129 Torino, Italy

Corresponding author: Francesco Ponzio (francesco.ponzio@polito.it)

This work was supported in part by the San Paolo Foundation under Grant CSTO164044, and in part by the Department of Neuroscience, University of Turin under Gran ROAS_AUTOF_18_01.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of University of Turin under Protocol No. 256076, and performed in line Declaration of Helsinki.

ABSTRACT So far, very little attention has been paid to the role of the autonomic nervous system in augmentative alternative communication solutions. In this regard, the pupil near reflex, one component of the triadic accommodative response to a visual plane shift in-depth, may play a key role. Such reflex does not necessitate any requirement of skeletal muscles, and thus may be preserved in diseases affecting somatic motoneurons, such as the amyotrophic lateral sclerosis. On this basis, the pupillary accommodative response, i.e. the pupil constriction in response to a *far-to-near visual shift in depth*, can constitute an effective communication tool, bypassing the well-known limitations of canonical eye-trackers, which require recruitment of skeletal muscles. This paper introduces e-Pupil: a low-cost, stand-alone, portable and smart device exploiting the pupillary accommodative response as a communication tool. The Internet of Things plays a crucial role in our device, guaranteeing portability, accessibility and usability, as well as several remote functionalities. We propose two different routes to communicate with the external world, both built upon the identification of the pupil constriction event in response to a voluntary far-to-near visual shift. The experimental validation proves the reliability of the system as well as its intrinsic simplicity of use.

INDEX TERMS Augmentative alternative communication, Internet-of-Things, pupillometry, image segmentation, complete *locked-in* syndromes.

I. INTRODUCTION

Communication is one of the most important abilities that humans have learned. It enables us to express our problems, feelings and needs as well as our thoughts and experiences. For this reason, communication problems, which connote several pathological situations, may severely affect life quality [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Md Kafiul Islam.

In extreme conditions, e.g., the so-called complete *locked-in* syndromes (CLIS), the impossibility to communicate is considered co-responsible for the rapid cognitive decline [2]. This condition may take place, for instance, in the final stages of amyotrophic lateral sclerosis (ALS), when paralysis eventually affects all skeletal muscles, thus preventing the use of traditional augmentative alternative communication (AAC) that exploits residual voluntary movements.

As a matter of fact, different types of assistive devices may exploit the residual activity in skeletal muscles for

communication purposes [3]. Extra ocular muscles are relatively spared by the initial phases of ALS and, as long as they remain functional, eye-tracking based AACs may offer good interaction possibilities including access to computers, internet and speech synthesizers, and control over domestic devices [4]. When progressive weakness or ocular dysfunction eventually makes the use of eye-trackers unfeasible, alternate solutions may consist in ad-hoc detection of residual muscle activity [5], [6] or brain-computer interfaces (BCI), mostly based on detection and interpretation of EEG signals [2], [3], [6], [7], [8]. These are, however, affected by a number of limitations including high cost, difficult set-up, requirement of long training periods, etc. [3].

The possibility to communicate through changes in pupil size has also been recently proposed [9], [10], [11], [12]. In particular, it was previously shown that, according to the so-called pupil accommodative response (PAR), pupil constrictions may be voluntarily produced by changing the focus from a far to a near visual target, and such constrictions may be easily detected and characterized [11]. This modality is of particular interest for AAC, given that the pupil size is governed by smooth muscles controlled by the autonomic nervous system [13], which may be preserved in pathological conditions affecting somatic motoneurons such as ALS [14], [15]. Moreover, the PAR may be voluntarily and consistently triggered by a subject by implementing a very natural task that does not require any long learning or training phase.

II. AUGMENTATIVE ALTERNATIVE COMMUNICATION: CURRENT PANORAMA

How important is communication to you? "*There are no words to describe how important it is for me to communicate independently! Personally, I can't imagine life without a communication device. It is as important for me as breathing and nutrition.*" —Rich, individual with late stage ALS, McLaughlin et al. [16].

Several neurological disorders may limit both verbal and nonverbal communication. Patients who are conscious and aware, but unable to communicate neither vocally nor by writing, are defined as *locked-in* subjects. In the "classic" locked-in syndrome (LIS), eye gaze movements and eyeblinks are preserved, and thus can be exploited as communication channel [17]. By contrast, in the *complete* locked in syndrome (CLIS) patients show also oculomotor impairment, and hence no communication is possible by any mean [17]. The impossibility to communicate emotions, feelings, thoughts, and needs is one of the most demoralizing difficulties for these patients [17]. As a consequence, the role of AAC facilitators is continuously gaining importance in a plethora of different communication impairments, going from stroke to neurodegenerative diseases, such as ALS [1]. Specifically to ALS disease, in a significant longitudinal review, Beukelman and colleagues observed that 96% of people with ALS for whom AAC was recommended, accepted and used AAC [1]. Furthermore, the authors also report that ALS patients who refused AAC demonstrated a co-occurring dementia or experienced multiple severe health issues concomitant with ALS [1]. This observation supports the hypothesis that the inability to communicate is co-responsible for the CLIS subjects' rapid cognitive decline [1], [2], and sets off the paramount importance of adequate AAC solutions for patients with total paralysis.

According to the American Speech Language and Hearing Association [18], an AAC system is "an integrated group of components used to enhance communication". Following this definition, AAC devices may be broadly categorized into two main groups: unaided or aided systems. Unaided forms of AAC, also refereed to as *No-Tech* (NT) solutions, do not need any external tools, but require a large degree of user's motor control. Gestures, manual signs, facial expressions, vocalizations, and body language are some examples.

Conversely, aided AAC systems require some form of external tools, and hence they are further divided into two sub-categories depending on the backbone support: (i) non-electronic based AAC solutions, often referred to as *Light-Tech* (LT) tools; (ii) electronic-based AAC solutions, also known as *High-Tech* (HT).

Communication boards are a prime example of LT technologies, typically consisting in a digital touchscreen depicting letters to compose words or symbols associated to a given meaning. On one hand using such devices does not require any specific training, on the other hand it involves a voluntary physical movement, which is not feasible for a wide range of subjects (e.g. LIS/CLIS subjects) [19].

Eye trackers are a class of HT devices which can be controlled leveraging the subject's eye movements. Motor neurons of the oculomotor, trochlear and abducens nuclei, which innervate the extrinsic muscles of the eye, controlling eye movements, are generally less prone to degeneration compared to upper and lower motor neurons of other voluntary muscles [20]. Based upon this observation, eye tracking-based technologies have been broadly used with LIS subjects [16]. Eye tracking-based AAC gathers a wide group of diverse technologies, ranging from digitized speech devices (for example, the *Megabee^{TM} Eye Pointing Communication Tablet* [21]) to computer-based speech generating devices built on top of commercial eye-trackers (for example, *Tobii EyeMobile Plus* [22]). Nonetheless, for a correct and productive use of the device, the user has to perform very fine eye movements to select specific items on the screen, which typically requires extensive training [3], [16]. Moreover such visual tasks are completely impossible for late stage LIS or CLIS subjects, due the progressive impairment of the visual skills associated with the neurodegenerative disease [16].

Switch scanning AAC technologies are another class of HT AAC devices, where items from a selection set are either highlighted by a scanning indicator (visual scanning) or announced via voice (auditory scanning). The user makes a selection by means of a switch, typically leveraging a minimal movement (e.g., jaw protrusion, thumb movement, etc.). Switch options currently available include mechanical switches, as well switches triggered by

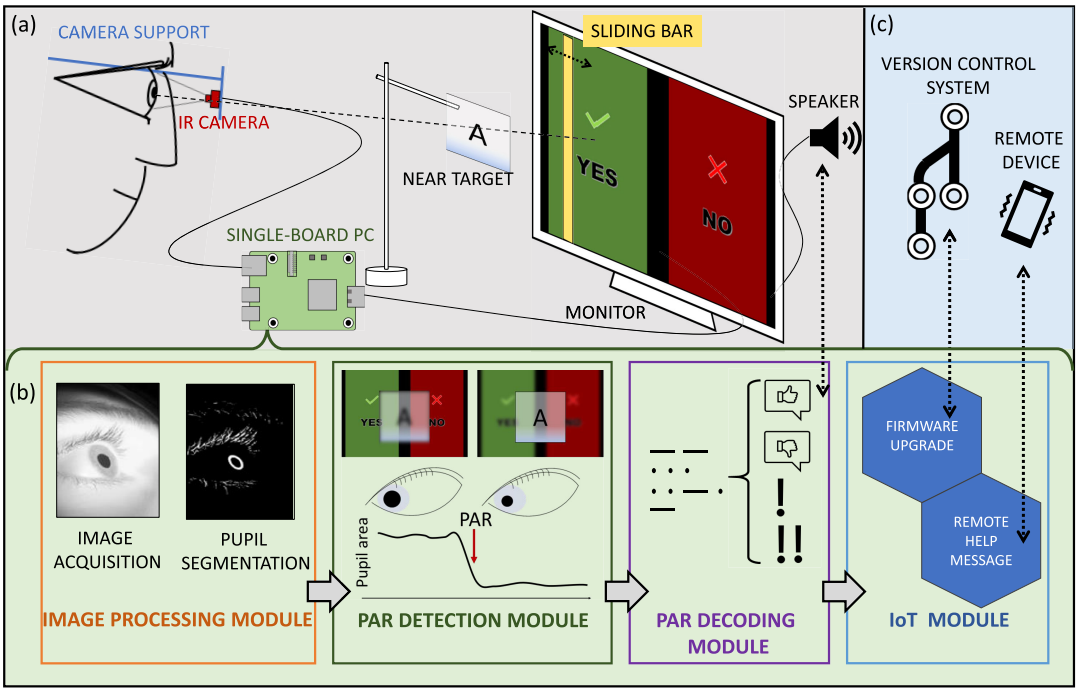


FIGURE 1. Schematic representation of the proposed device. (a) hardware set-up; (b) software modules implemented on the single board computer; (c) remote components that can communicate with the device via internet.

electromyography (EMG), electrooculography (EOG), proximity sensors, or fiber-optic sensors [16]. Given the progressive limitation of their movement capability, the nature and sensitivity of the switch mechanism is especially critical for the LIS/CLIS users.

Additional typical issues such as poor switch placement, unintentional switch triggering and difficult mounting are also a cause of frustration for most of these subjects, which often leads to the abandonment of the device [16].

Finally, the most frontier class of HT ACC technologies are the *Brain Computer Interfaces (BCIs)* leveraging the electroencephalographic (EEG) signal. Brain signals for BCI may be acquired either invasively or noninvasively. Invasive BCI systems require to surgically implant an electrode array directly onto the cerebral cortex, while noninvasive systems typically rely on electrodes placed against the scalp [16]. The EEG signal can be exploited as the basis of different types of communication protocols. For example, *Sensorymotor Rhythms-based* BCIs [23] encode the subject's *intention* to move some specific muscles into a message: e.g., intention to move the right hand is "yes", intention to move the left hand is "no". As BCIs do not require any voluntary movements of the subject, they may be the ideal solution even in the last stages of neurodegenerative disorders such as the ALS disease. Nonetheless, fatigue, cognitive changes and visual impairments that are typically associated with ALS LIS/CLIS are known to affect BCI performance [16]. Moreover, they are also affected by environmental factors such as ambient noise and interference with other medical equipment (e.g. mechanical ventilation) [16]. These factors, added to the high costs of the equipment and the cumbersome procedures for preparing the device (e.g., correctly positioning the electrodes on the scalp) and training the subject, still limit the widespread diffusion of such devices in the standard practice [24].

In general, a key aspect for a larger adoption of AAC solutions in LIS/CLIS cases is the need of more flexible low-cost systems, able to adapt to the special needs of the end-users without the necessity of expensive and/or complex set-up. Starting from such considerations, in this paper we present *e-Pupil*, a PAR-based AAC system. Our solution tries to address the limitations of current AAC devices, by featuring: (i) easy management of the basic communication tasks based on PAR, as later detailed, which does not require any cumbersome training of the subject [11]; (ii) smart IoT features to allow remote emergency calls, as well as continuous remote access for patient monitoring and automatic software upgrade; (iii) low-cost hardware/software equipment and maintenance.

III. EXPLOITING PAR AS AAC SOLUTION

Whenever we look at objects located at different depth planes with respect to the optical axis, the pupil size changes in a coordinated way, with lens accommodation and ocular convergence/divergence according to the so-called accommodation reflex. In particular, the pupil constricts in response to a far-to-near shift and dilates in response to a near-to-far shift. Note that, although the current pupil size is controlled by the autonomic nervous system, it can indirectly be triggered by the voluntarily decision to focus on a near or far visual target, and hence may serve as the basis for AAC [11].

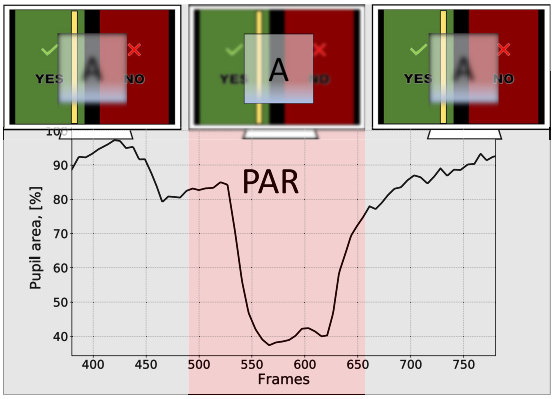


FIGURE 2. Representative example of PAR signal. The pupil constricts in response to a *far-to-near* gaze shift, when the subject focuses on the *near* target (A letter).

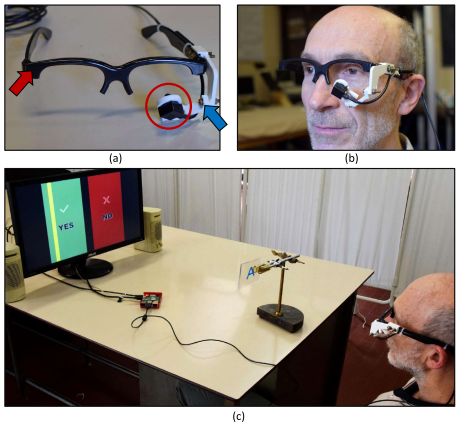


FIGURE 3. Photographs of hardware components of e-Pupil. (a) Infrared camera (red circle) attached to adjustable holder (blue arrow) mounted on a spectacle frame (red arrow). (b) The small infrared camera focuses on the left eye of the subject. (c) Complete set-up of the prototype.

FIGURE 1(a) shows a scheme of our solution. In the baseline condition, the subject stares at a monitor placed at a distance of about 2 meters. The screen presents a graphical user interface showing two sectors, respectively with green and red background, and a yellow sliding bar in foreground, which cyclically crosses the whole screen from side to side (more details will follow). A 5 × 5 cm piece of transparent plastic, with a small *A* letter written on it, is positioned between the screen and the subject's dominant eye through a holder. In our setting, the screen serves as *far target* and the *A* as *near target*: as long as the subject points the gaze steadily to the screen, the pupil maintains a constant size. When the subject performs a visual plane change by shifting the focus on the *near target* (i.e., the *A*), the pupil constricts.

FIGURE 2(b) bottom shows a clear example of the signal that can be obtained by monitoring the dominant eye of a subject performing a PAR. When the subject voluntarily performs the far-to-near gaze shift, the area of the pupil undergoes a sudden significant reduction. *e-Pupil* aims to exploit this effect to perform a selection through the sliding bar, which, in the simplest case will return a binary *Yes/No* answer. However, additional features may be extracted, as later explained.

IV. SYSTEM ARCHITECTURE

FIGURE 1 schematically represents the architecture of *e-Pupil*. The different sections of the figure provide: (a) the hardware set-up of the system; (b) the software modules implemented on the single board computer (SBC); (c) the external components that can communicate via internet with the device, to put its remote functionalities into effect. The hardware components of the real prototype are shown in FIGURE 3. They consist of:

1) *Single-board computer*. We intended to design a low cost, portable and flexible device. Hence, the choice of a computer desktop or laptop was not practicable. In this respect, a SBC is ideal to minimize the price-performance ratio. For this purpose, we chose the *Raspberry Pi 3 model B* board. Our preliminary experiments involving image acquisition and basic processing demonstrated that the computational power of the *Raspberry Pi 3* board is enough for the purpose of our application by still maintaining the overall cost reasonably low (less than 50 USD). Nonetheless, the system can be easily modified to host a different SBC.
2) *Infrared camera*. We used a low-cost, 2MP infrared camera by Shenzhen DingDang Smart Technology CO., LTD with resolution 1080 × 1080 pixels, which is fully compatible with the Raspberry Pi through a USB connection (see red circle in FIGURE 3(a)).
3) *Camera support*. For a correct and stable image acquisition, the infrared camera is fixed to a spectacle frame (red arrow in FIGURE 3(a)) by means of a 3D-printed adjustable holder (about 10 cm long), providing for optimal camera positioning and orientation through 4 degrees of freedom (blue arrow in FIGURE 3(a)).
4) *Display monitor*. An *Asus VW193S* Monitor (21,5 inches dimension) is connected via High-Definition Multimedia Interface (HDMI) to the SBC.
5) *Speakers*. Two *juster AC691N* speakers are connected via the SBC headphone jack.

The software modules, all implemented in *Python*, are: the *Image Processing*, *PAR Detection*, *PAR Decoding* and the *IoT module*, respectively.

FIGURE 1(b) provides an overview of their functionality and of the way they are interconnected. The input of the processing pipeline is a single frame acquired by the IR camera, which typically includes not only the pupil of the subject but also the iris, the eyelashes and a portion of the subject's face. The *Image Processing* module identifies the area of the pupil in each input frame, by applying image segmentation algorithms. Then, the signal constituted by the area of the subsequent frames is fed into the *PAR Detection* module. If a PAR is detected, the *PAR Decoding* module is in charge of its interpretation, that is, of activating and sending the corresponding audio signal to the speakers. More specifically,

based on the characteristics of the PAR event, the PAR Decoding module may activate one of the following audio signals:

(i) yes/no answers (represented by thumbs-up and down, in FIGURE 1(b));
(ii) a message summoning the attention of the other person (represented by the single exclamation mark);
(iii) a message requesting immediate and urgent help (double exclamation mark, in the figure).

In the latter case, the help message is also sent remotely via the *IoT module* (for example, to the caregiver's smartphone). This module is also in charge of the automated firmware update, that is handled by connecting to a remote version control system (VCS).

In the following, we go more in-depth into the implementation of the individual modules. As far as the IoT functionalities are concerned, they will be discussed later in the dedicated Section V.

A. IMAGE PROCESSING MODULE

Through this module, a single frame is initially acquired from the IR camera via *OpenCV* functionalities [25], and then fed into an automated image segmentation pipeline, aimed at extracting the area of the pupil. The acquired RGB frame is first converted to gray-scale and then cropped into a smaller size of 320×480 pixels. This ensures a better processing speed-up, coupled with a reasonable resolution for the downstream image elaboration.

The cropped frame undergoes a *background subtraction* phase, in which the background, obtained via a median filter with kernel size 21, is removed from the original image in order to enhance the contrast of the pupil with respect to the rest of the image. Then, a global *intensity thresholding* leveraging Otsu algorithm [26] provides the binary mask of the pupil, together with some other smaller spurious connected regions (typically, parts of the eyelashes and/or of the iris). To regularize the segmented shapes, a morphological dilatation operator is applied through a 5×5 pixels kernel, followed by a flood-fill algorithm. This step ensures the compactness of the pupil's mask and makes it identifiable against all the other spurious regions.

Stemming from the consideration that the pupil should be the largest and roundest connected region in the binary mask, the following steps are implemented:

(i) perimeter (P), area (A), and roundness ($R = \frac{4\pi A}{P^2}$) are computed for each connected region;
(ii) regions with area below threshold $\theta = CS * 0.2$ are discarded. Here, CS is the size of the crop. Thus, θ is proportional to the resolution of the camera;
(iii) the individual region with highest R value is recognized as the pupil.

Then, the area of the pupil (that is, the number of pixels of the corresponding mask) is fed into the *PAR Detection* module, meaning that a PAR event will be solely detected based on the change of area of the pupil's region. The reason behind choosing the area over other figures of merit (for example radius) is that the area scales with the square of the radius, which makes changes due to dilatation/constriction events more evident.

Overall, the overhead due to the image processing pipeline limits the available frame-rate to 11*fps* for downstream analysis. In our experiments, we found that this frame-rate is more than enough to detect PAR events, even in experimental settings where the frequency of pupil's constrictions had been pushed to its physiological limits [11].

B. PAR DETECTION MODULE

FIGURE 2 shows a representative example of the time course followed by the pupil size when a subject performs a far-to-near visual shift, thus inducing a PAR. As it can be gathered from the plot, even though some physiological fluctuations exist, the PAR is clearly identifiable. Given a generic i^{th} frame f and the corresponding area value $A_i(f)$, to detect a PAR we proceed as follows:

(i) to ensure robustness over signal's fluctuations, we compute the mean value of the area within a sliding window, which includes the last $N = 2 \cdot framerate$ previous samples:

$$\mu = \frac{1}{N} \sum_{i=i-N}^{i} A_i(f); \tag{1}$$

(ii) $A_i(f)$ is identified as PAR candidate if it satisfies the following criterion:

$$A_i(f) < \alpha \cdot \mu \tag{2}$$

with $\alpha = 0.85$, empirically set.
(iii) if inequality (2) is satisfied for 4 consecutive area samples, a corresponding PAR event is detected, and the downstream blocks are triggered.

C. PAR DECODING MODULE

As already introduced in Section III, our device allows the subject to interact with another person (e.g., a caregiver) either in an *answering* or in a *calling* mode.

(i) *Answering mode*.
This modality allows the subject to provide yes/no answers to direct questions made by the other person. As it can be gathered from FIGURE 2 top, the graphical user interface of *e-Pupil* consists of two main zones, respectively colored in green (left) and red (right). Intuitively, the green and the red regions are respectively assigned to *"Yes"* and *"No"* answers. A yellow sliding bar in the foreground (see FIGURE 2 top) continuously crosses the screen from left to right in a cyclical way. Suppose the subject is asked a question. If the subject intentionally shifts the focus on the A (and hence, triggers a PAR event) while the sliding bar is passing over the green region, the answer is taken to be a *"Yes"* (see FIGURE 2 top-center). The system will react by playing a recorded voice, saying *"Yes"*. Accordingly, to say a *"No"*, the subject should trigger the PAR event while

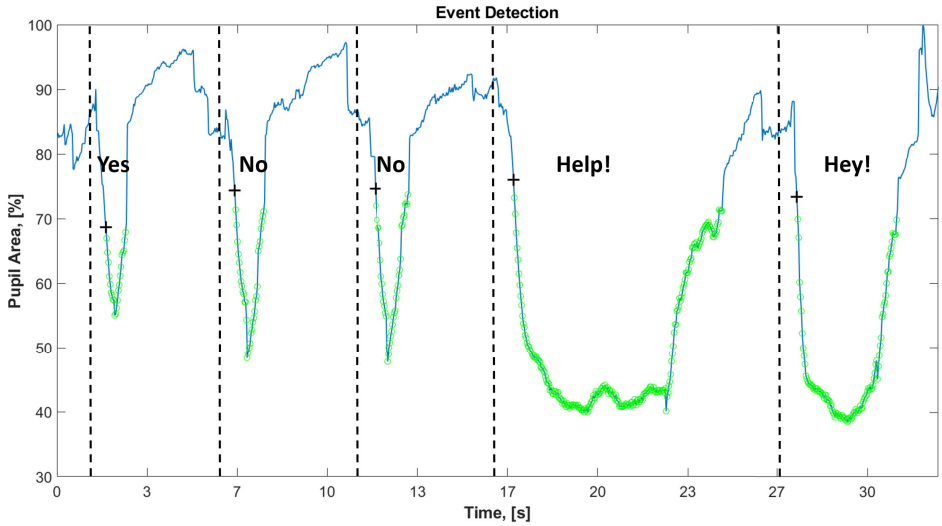


FIGURE 4. Representative example of the communication routes exploited by the proposed device. The green tracing identifies the PAR detection and duration: to rise a request of help the constriction must be maintained for at least 6 seconds.

the yellow bar is passing inside the red zone. If the PAR is triggered in an ambiguous zone, namely when the sliding bar is crossing the black area between the *"Yes"* and *"No"* regions or in the zones located at the border of the screen, the system will play a recorded message expressing uncertainty. To ensure a pleasant animation, two parallel threads take charge respectively of the image processing pipeline and of the animation of the interface, and communicate with each other. At the exact moment the subject executes a pupil constriction task, the position of the sliding bar is identified, allowing the system to interpret the subject's response and to act accordingly.

(ii) *Calling mode*.
While in answering mode the conversation is initiated by the operator/caregiver, in the calling mode the subject/patient independently activates the call. This communication modality gives the opportunity to actively summon somebody's attention or even ask for urgent help, in case of necessity.
In this case, the solution leverages on the duration of the PAR event, independently of the position of the sliding bar. More specifically:

a) if the subject induces a PAR lasting more than 6 seconds, the system activates a urgent distress call. This triggers two different events: a local audio message (see *Help*, in FIGURE 4), as well as a remote request via internet (for example, with a *Telegram bot*). This latter feature is described in more detail in the following section, where the IoT functionalities of the device are discussed;
b) if the subject produces a PAR that lasts between 3 and 6 seconds, no matter the position of the yellow sliding bar, the system activates a recorded audio summoning attention (*Hey* message, in FIGURE 4).

V. DISTRIBUTED SOFTWARE PLATFORM FOR IoT DEVICE MANAGEMENT

As already anticipated, *e-Pupil* is an IoT device allowing remote features, such as: i) emergency calls, ii) access to recorded data for patient monitoring and iii) automatic software upgrade. For these purposes, we adopted and extended *iottly*, which is a distributed software platform to manage IoT devices [27]. It enables data transport between devices and applications over secure connections by exploiting two communication paradigms, namely the request/response through REST [28] and the publish/subscribe [29] through MQTT [30]. REST is a synchronous architectural style used to build web services exploiting the HTTP protocol. It is widely supported and suited for 1-to-1 communications. On the contrary, MQTT is an asynchronous lightweight protocol optimal for 1-to-many and (near-) real-time event-driven communications. Hence, all of the IoT devices implement unified communication interfaces independently on the specific functions they are running, easing data transmission and device management. Moreover, devices are mutually authenticated as their certificates (2048 bit) are signed by an internal Certification Authority (CA), thus following the X.509 standard. This allows confidentiality, integrity, authenticity and non-repudiation of the information transmitted between the different actors, either hardware or software, in the platform.

The main features provided for device control and management are:

- remotely reboot of devices to recover from software issues or unexpected conditions;

- remotely restart of specific processes to restore their original conditions;
- monitor critical processes by receiving (near-) real-time notifications about their running status;
- transfer diagnostic logs by timely obtaining valuable information about the device operating status;
- transfer files from/to devices;
- enable remote access to devices via ssh sessions.

In its core, the platform consists of two main macro-components: i) *iottlyCloud* and ii) *iottlyAgent*. As reported in FIGURE 5, *iottlyCloud* allows the distributed software components to be deployed into cloud systems as Docker containers. Such components are needed to manage i) authentication of devices (i.e., *Internal CA* and *Authentication* modules), ii) remote ssh connections (i.e., *SSH* macro module and all its sub-components) and iii) *Database Replicas*. To enable MQTT connections, *iottlyCloud* provides a *Message Broker*, which is a server that gathers all messages from a given number of clients (a.k.a publishers), and then routes such messages to the appropriate destination clients (a.k.a subscribers). The information flow is arranged in a hierarchy of topics: when a publisher has a new instance of data to send, it transmits a message with the data to the connected Message Broker with a specific topic. The Message Broker then distributes the information to any subscribers that have registered to that specific topic. The publisher is totally unaware of the number or the locations of subscribers, and subscribers, on their part, do not have to be configured with any data about the publishers. The *Software Repository* module provides the latest version for the software algorithms (i.e. *Image Processing*, *PAR Detection* and *PAR Decoding*) to be downloaded by *e-Pupil* devices via REST. To make the whole system scalable, different replicas of the same software component can be deployed and managed by a *Load Balancer*. Finally, specific REST API (i.e., API servers module) are provided by iottlyCloud to automate and integrate the management of the devices with third-party applications.

The *iottlyAgent* in FIGURE 6 is a software component to be deployed on the edge, which can coexist without interfering with third-party firmware. It is based on static *Python* tailored for embedded devices and is in charge of integrating the IoT devices in our infrastructure. It provides specific features to execute and manage on edge either native scripts developed with the *iottly SDK* (i.e., *SDK Manager*, *Script engine*, *Script OTA sync* and *Script execution* modules in FIGURE 6) or *Third-party software*. To allow data messaging in (near-) real-time, it offers an *MQTT client*, which works either as publisher and/or subscriber. The *SSH manager* provides functionalities to run both an *SSH server* and *SSH clients* for a direct remote management of on edge devices.

As mentioned in Section IV, *e-Pupil* is based on a Raspberry Pi 3 model B, which runs an instance of the *iottlyAgent* together with our algorithms for *Image Processing*, *PAR Detection* and *PAR Decoding*. Secure and timely remote updates are essential to have a device always working and totally reliable, possibly reducing the cost of maintenance of the hardware, that is naturally located at a subject's home. Thanks to *iottlyAgent* features, a background process periodically checks the software version installed on the *e-Pupil* device, comparing it with the last updated code version, stored in the remote *Software Repository* module. In case the version id of the local software is different from the one in the repository, an automatic download and install is triggered. This ensures that *e-Pupil* devices are always updated to the latest software version, without needing any operator's intervention.

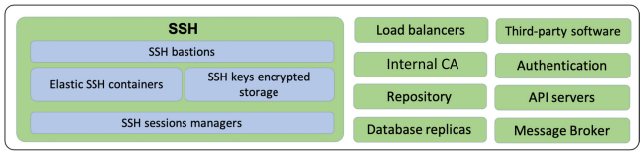


FIGURE 5. ***iottlyCloud*** **architecture.**

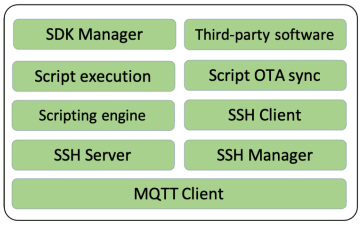


FIGURE 6. ***iottlyAgent*** **architecture.**

As discussed in Section IV, our *e-Pupil* IoT device leverages a speaker to make the user able to communicate with the outside world. To avoid possible unreported alarms caused by hardware related problems (e.g., volume too low) or temporary disattention of the caregiver, the system activates a remote distress call sent in (near-) real-time via MQTT. In our setting, each *e-Pupil* device acts as publisher, sending possible distress call events on a specific topic. We chose to implement a *Telegram Bot* [31] as subscriber, constantly waiting for messages from the *e-Pupil* devices.

The choice of integrating the Telegram between our platform and the final destination clients is two-fold. On the one hand, we want a messaging interface as much user-friendly as possible. In this sense, Telegram Bots are ideal, as the users can interact with them by sending messages, commands and inline requests in an intuitive chat-style fashion (see an example in FIGURE 7). On the other hand, we prove the integration of a remote third-party software into our distributed software platform testing their interoperability.

Once correctly connected on Telegram, the caregiver can check the notification state of the remote distress calls, and is always reachable by *e-Pupil*'s user. Thus, in the very moment when a subject arises a distress call, *e-Pupil* publishes a message, which is routed by the Message Broker and, then, the subscriber (i.e., our Telegram Bot) sends an alarm notification to the destination mobile client associated with the given patient, through the specific chat-id assigned by Telegram.

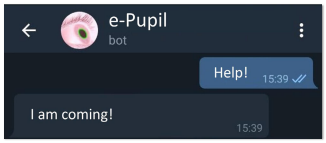


FIGURE 7. Chat-style fashion ensures a comfortable and intuitive user-friendly interface.

VI. EXPERIMENTAL CHARACTERIZATION

In this Section, we assess the performance of our system, by evaluating independently: (i) the performance of the image processing pipeline, in terms of pupil's segmentation quality; (ii) the performance of PAR events detection; (iii) the accuracy of the PAR decoding, in terms of correct response of the system to yes/no/attention/help requests made by the subject.

A. IMAGE PROCESSING ACCURACY

Since the variation of the pupil area is the backbone of PAR events detection, the quality of the segmentation of the pupil impacts all the downstream tasks. To assess this quality, we collected recordings from twenty different healthy subjects (11 males and 9 females) recruited from the University of Turin, aged between 23 and 61, who tested our device. To test the robustness of the segmentation pipeline, such recordings were performed with different ambient light conditions, ranging from 600 to 800 lux, measured by means of a luxmeter (Testotermo 0500, Dott. Ing. S. Ciano, Italy, Torino). As a quantitative measure of the segmentation quality, we adopted the Dice coefficient over 200 frames manually segmented as ground truth. Such index is a metric used to evaluate the overlap between two discrete sets (X and Y), as follows:

$$\frac{2|X \cap Y|}{|X| + |Y|}$$

The Dice coefficient is commonly employed to evaluate the similarity between binary images, which in our case correspond to the ground-truth mask of the pupil, manually segmented in each frame, and the corresponding mask obtained by our segmentation pipeline. Overall, we obtained a mean Dice index of 0.89 ± 0.03 (mean value $\pm$ standard deviation), which demonstrates a very good correspondence between automated segmentation and manual ground-truth.

B. PAR DETECTION ACCURACY

As extensively discussed in the previous Section, both the communication modalities exploited by our device depend on the correct detection of a given PAR event.

To assess the trustworthiness of our system in detecting such events, we used recordings from the same twenty subjects of Section VI-A, and put in place the following validation procedure, same as in [11]. The subjects were given auditory cues at two different tones (high-frequency and low-frequency, respectively at 2000 and 250 Hz), and asked to shift the focus to the near target only in response to the high-frequency cues. In all the other cases (i.e., during the low-frequency clues, as well as during silent intervals), they were instructed to maintain the focus on the far target. A randomized sequences of 10 audio cues (5 at low and 5 at high frequency) were presented, at the rate of 15 cues/min. The duration of a single cue, as well as the inter-cue interval, was 2 s. A 2-min resting interval was always inserted between two following sequences, which are presented in random order. To avoid premature responses, possibly triggered by expectation, a jitter of $\pm$ 0.25 s was randomly added to the inter-cue resting interval.

Then, the PAR detection procedure described in Section IV was exploited to detect PAR occurrences correspondent to the given audio cues. Thus, was compared to the audio cue sequence, as follows: a positive score was assigned whenever a high-frequency cue matched a PAR or a low-frequency cue matched a no-PAR; a negative score was given otherwise.

Then, we assessed the performance of PAR detection with two different figures of merit:

(i) accuracy (A) of the information transfer, that is computed as the percentage of correctly decoded answers (i.e., number of positive scores divided by the total number of scores)

(ii) rate (R) of the information transfer, that is quantified as in [7]:

$$R = V * (\log_2 N + A * \log_2 A + (1 - A) * log_2(\frac{1 - A}{N - 1}))$$

where V is the rate at which the audio cues were presented and $N = 2$, for binary communication [7].

In our experiments, we obtained very satisfactory numbers for both the figures of merit: the accuracy of information transfer was 0.99 ± 0.01 (mean value $\pm$ standard deviation), coupled with an information transfer rate of 13.8 bits/min.

C. PAR DECODING ACCURACY

To assess the accuracy of the system in translating the PAR events into the proper output messages, we again used the recordings from the twenty volunteers. Each subject was asked to voluntarily provide yes/no answers, as well as to elicit requests of attention (i.e., with PAR maintained for more than 3 seconds) and of urgent help (i.e., PAR duration above 6 seconds). The order of the yes/no answers as well as of the attention/help requests (four in total, per each subject) was randomized.

At the end of the experiment, we compared the answers triggered by the communication device with the expected ones, by interpreting the four different types of outcomes (yes, no, attention and help requests, respectively) as labels of a classification task with four different classes. Thus, the results of the comparison can then be reported in the form of a confusion matrix (see FIGURE 8), where the true and the predicted classes correspond to the expected and the automatic outcomes, respectively. As it can be gathered from the figure, the system was 100% accurate for all four categories, and none of the requests was misinterpreted.

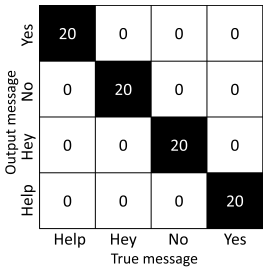


FIGURE 8. Confusion matrix of the PAR decoding task.

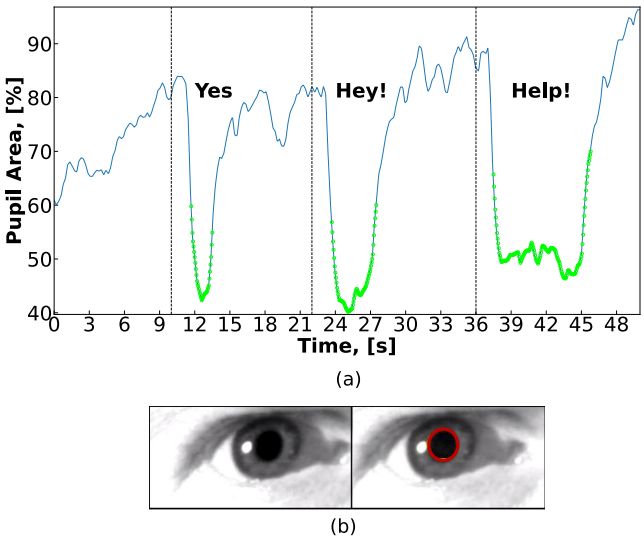


FIGURE 9. Results obtained from the trial on a late-stage LIS ALS patient. (a) Communication routes retrieved from elicited PAR events leveraging *e-Pupil*. The green tracing identifies the PAR detection and duration. Answering mode (*Yes*) as well as calling mode (*Hey!* and *Help!*) are proficiently evoked by the subject and detected and interpreted by our device. (b) Patient's pupil segmentation (right) beside the original image (left).

VII. EXPERIMENTAL VALIDATION ON ALS PATIENT

The device was also tested by a bed-ridden, late stage ALS patient (male, 63 years old). At the time of the experimental session, the patient presented complete tetraplegia and anarthria and was artificially ventilated. He was however still able to use an eye-tracking based communicator. A set-up including far and near target, as depicted in Figure 1(a), was arranged and the patient was instructed to shift the focus from far to near, and to maintain it on the near target for different time lengths adequate to activate the different responses from the system, namely, the selection of the *Yes/No* answer (1s $<$ duration $<$ 3s), the request of attention (3s $<$ duration $<$ 6s) and the request of help (duration $>$ 6s). The patient had no difficulty in understanding the task and managed to evoke the expected PAR successfully. Original recordings of pupil size are presented in Figure 9(a). These results, besides supporting the feasibility of exploiting PAR in ALS patients [15], document the easiness for the patient to produce clear-cut responses on the first experimental session, as well as of controlling the duration of the myosis.

VIII. DISCUSSION

As mentioned in the literature review, pupil-based communication has been scarcely taken into consideration in the AAC context. Most studies focused on pupil size variations evoked by changes in emotional states and cognitive efforts [32]. The preliminary study by Ekman et al. [33] discussed several ways to voluntary modify the pupil size: physical and mental efforts, change in point of focus (which is equivalent to PAR), positive and negative emotions and concentration. Nonetheless, the results of a successive study were discouraging to the authors [34], and not followed-up. Thus, this issue remained mostly unknown to the biomedical community. On the other hand, the large evolution of eye-tracking devices has been limited to screen-based applications, where the depth of focus never changes, which definitively precludes any possible manifestation of the PAR. The main goal of our current study was to describe and characterize the first prototype of *e-Pupil*: a low-cost, portable and reliable AAC device, built on top of the detection of a PAR event in response to a far-to-near shift of focus, which is voluntarily triggered by the user. Our previous study demonstrated that PAR can be effectively exploited as AAC, thanks to its inherent stability, high transfer rate, and "intuitive" nature [11]. Of course, pupil size is also affected by changes in ambient luminance and sudden emotional stimuli, which may potentially disturb the measurement. However, ALS patients normally live in quite and stable environments in which major changes in ambient luminance are not expected to occur or could be easily prevented. As for emotional stimuli, they would mostly evoke pupil dilations (mediated by the sypathetic nervous system), rather then sharp constriction, and thus should be clearly distinguishable from PARs and not misinterpretated. It is also worth to emphasize that, differently from other brain-computer interfaces, eliciting a PAR does not necessitate to learn complex procedures or demanding mental tasks, and is a viable solution in conditions that prevent the use of standard eye-tracking-based AACs, such as the presence of nystagmus [35].

As compared to previous implementations [11], [35], the prototype introduced in the present study represents a relevant step forward for the exploitation of the PAR as a mean of communication because i) it is a compact, low-cost, and stand-alone device, which can be autonomously handled and used by patients and caregivers without the presence of expert personnel; ii) the patient performance and actual pupil size changes are recorded and can be remotely controlled; iii) in addition to the simple PAR detection, the consideration of PAR duration, as an additional feature of the PAR event, has been successfully implemented. Indeed, the results showed that the device is accurate in translating the PAR events into yes/no answers, as well as in the management of attention/help requests, even remotely. Our *e-Pupil*'s performance appears to be considerably high, when compared to other AAC communication systems. Specifically, the results showed a 99% accuracy at 13.8 binary cues/min, which is significantly higher with respect to 90% of accuracy at

4.86 binary selections/min of the communication based on attentional modulation of pupil size [9]. Lastly, as described in section VII, our device was also tested by a bed-ridden, late stage ALS patient. The subject, who was asked to elicit a yes answer, a requests of attention, and of urgent help, was able to successfully communicate his intentions leveraging *e-Pupil* features. Furthermore, the patient was immediately able to proficiently use the device, suggesting a reduced cognitive load for learning the communication system, which corroborates the "intuitive" nature of the proposed AAC solution. Although no general conclusion can be drawn from a single patient, these preliminary results are very promising and support the concept that PAR is a viable communicative pathway in ALS.

IX. CONCLUSION

In conclusion, the proposed system allows an accurate communication without the necessity of any movement. Unlike BCI systems, it does not require any particular training for the user, furthermore its cost is much lower. The device is enriched with some useful IoT features. One of them makes the prototype a reliable system which is always updated. The other allows the user to receive immediate help, contacting his/her caregiver through remote. Further studies need to be performed to fully verify the usability of the proposed system for ALS patients both in locked-in state and not, which we soon plan to pursue within a clinical trial.

REFERENCES

[1] D. R. Beukelman, S. Fager, L. Ball, and A. Dietz, "AAC for adults with acquired neurological conditions: A review," *Augmentative Alternative Commun.*, vol. 23, no. 3, pp. 42–230, Sep. 2007. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17701742

[2] U. Chaudhary, N. Birbaumer, and A. Ramos-murguialday, "Brain-computer interfaces for communication and rehabilitation," *Nature Rev. Neurology*, vol. 12, no. 9, pp. 513–525, 2016.

[3] S. Koch Fager, M. Fried-Oken, T. Jakobs, and D. R. Beukelman, "New and emerging access technologies for adults with complex communication needs and severe motor impairments: State of the science," *Augmentative Alternative Commun.*, vol. 35, no. 1, pp. 13–25, Jan. 2019.

[4] A. Calvo, A. Chiò, E. Castellina, F. Corno, L. Farinetti, P. Ghiglione, V. Pasian, and A. Vignola, "Eye tracking impact on quality-of-life of ALS patients," in *Computers Helping People With Special Needs* (Lecture Notes in Computer Science), vol. 5105, K. Miesenberger, J. Klaus, W. Zagler, and A. Karshmer, Eds. Berlin, Germany: Springer, 2008, pp. 70–77. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-540-70540-6_9

[5] A. Tonin, A. Jaramillo-Gonzalez, A. Rana, M. Khalili-Ardali, N. Birbaumer, and U. Chaudhary, "Auditory electrooculogram-based communication system for ALS patients in transition from locked-in to complete locked-in state," *Sci. Rep.*, vol. 10, no. 1, pp. 1–10, May 2020.

[6] D. R. Beukelman, L. J. Ball, and S. Fager, "An AAC personnel framework: Adults with acquired complex communication needs," *Augmentative Alternative Commun.*, vol. 24, no. 3, pp. 67–255, 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18830913

[7] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, and T. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophys.*, vol. 113, no. 6, pp. 767–791, 2002.

[8] I. Lazarou, S. Nikolopoulos, P. C. Petrantonakis, I. Kompatsiaris, and M. Tsolaki, "EEG-based brain–computer interfaces for communication and rehabilitation of people with motor impairment: A novel approach of the 21st century," *Frontiers Human Neurosci.*, vol. 12, pp. 1–18, Jan. 2018.

[9] 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, pp. 1–15, 2016, doi: 10.1371/journal.pone.0148805.

[10] J. Stoll, C. Chatelle, O. Carter, C. Koch, S. Laureys, and W. Einhäuser, "Pupil responses allow communication in locked-in syndrome patients," *Current Biol.*, vol. 23, no. 15, pp. R647–R648, Aug. 2013, doi: 10.1016/j.cub.2013.06.011.

[11] F. Ponzio, A. E. L. Villalobos, L. Mesin, C. de'Sperati, and S. Roatta, "A human-computer interface based on the 'voluntary' pupil accommodative response," *Int. J. Hum.-Comput. Stud.*, vol. 126, pp. 53–63, Jun. 2019, doi: 10.1016/j.ijhcs.2019.02.002.

[12] C. de'Sperati, S. Roatta, N. Zovetti, and T. Baroni, "Decoding overt shifts of attention in depth through pupillary and cortical frequency tagging," *J. Neural Eng.*, vol. 18, no. 3, Jun. 2021, Art. no. 036008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/32348980

[13] W. Neuhuber and F. Schrödl, "Autonomic control of the eye and the iris," *Autonomic Neurosci.*, vol. 165, no. 1, pp. 67–79, Nov. 2011, doi: 10.1016/j.autneu.2010.10.004.

[14] R. Baltadzhieva, T. Gurevich, and A. D. Korczyn, "Autonomic impairment in amyotrophic lateral sclerosis," *Current Opinion Neurol.*, vol. 18, no. 5, pp. 487–493, 2005.

[15] S. Martinez-Conde, L. Martinez-Otero, A. Compte, and R. Groner, "Abstracts of the 20th European conference on eye movements, 18–22 August 2019, in Alicante (Spain)," *J. Eye Movement Res.*, vol. 12, no. 7, pp. 18–22, Nov. 2019.

[16] D. McLaughlin, B. Peters, K. McInturf, B. Eddy, M. Kinsella, A. Mooney, T. Deibert, K. Montgomery, and M. Fried-Oken, "Decision-making for access to AAC technologies in late stage ALS," *Augmentative Alternative Commun., Challenges Solutions*, pp. 169–199, 2021.

[17] A. Kübler, B. Kotchoubey, J. Kaiser, J. R. Wolpaw, and N. Birbaumer, "Brain–computer communication: Unlocking the locked in," *Psychol. Bull.*, vol. 127, no. 3, pp. 358–375, May 2001.

[18] *American Speech Language and Hearing Association*. Accessed: Jul. 15, 2021. [Online]. Available: https://www.asha.org/

[19] *Tactus Therapy AAC Alphabet Board*. Accessed: Jul. 15, 2021. [Online]. Available: https://tactustherapy.com/app/alphatopics-aac/

[20] J. Nijssen, L. H. Comley, and E. Hedlund, "Motor neuron vulnerability and resistance in amyotrophic lateral sclerosis," *Acta Neuropathologica*, vol. 133, no. 6, pp. 863–885, 2017.

[21] *Megabee*. Accessed: Jul. 15, 2021. [Online]. Available: https://www.megabee.net/

[22] *Tobii Eyemobile*. Accessed: Jul. 15, 2021. [Online]. Available: https://www.tobiidynavox.com/devices/eye-gaze-devices/eyemobile-plus-surface-pro/

[23] H. Yuan and B. He, "Brain–computer interfaces using sensorimotor rhythms: Current state and future perspectives," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1425–1435, May 2014.

[24] J. S. Brumberg, K. M. Pitt, A. Mantie-Kozlowski, and J. D. Burnison, "Brain-computer interfaces for augmentative and alternative communication: A tutorial," *Amer. J. Speech-Language Pathol.*, vol. 27, no. 1, pp. 1–12, 2018. [Online]. Available: https://pubs.asha.org/doi/abs/10.1044/2017_AJSLP-16-0244

[25] *Opencv*. Accessed: Jul. 15, 2021. [Online]. Available: https://opencv.org/

[26] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. SMCB-9, no. 1, pp. 62–66, Feb. 1979.

[27] *Iottly*. Accessed: Jun. 2021. [Online]. Available: https://iottly.com/

[28] R. T. Fielding and R. N. Taylor, "Principled design of the modern web architecture," *ACM Trans. Internet Technol.*, vol. 2, no. 2, pp. 115–150, May 2002.

[29] P. T. Eugster, P. A. Felber, R. Guerraoui, and A.-M. Kermarrec, "The many faces of publish/subscribe," *ACM Comput. Surv.*, vol. 35, no. 2, pp. 114–131, 2003, doi: 10.1145/857076.857078.

[30] *MQTT*. Accessed: Jun. 2021. [Online]. Available: http://mqtt.org

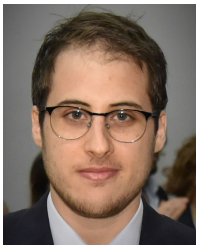
[31] *Telegram*. Accessed: Jul. 15, 2021. [Online]. Available: https://core.telegram.org/

[32] P. Majaranta and A. Bulling, *Advances in Physiological Computing*, S. H. Fairclough and K. Gilleade, Eds., 2014, pp. 17–39. [Online]. Available: http://link.springer.com/10.1007/978-1-4471-6392-3

[33] I. M. Ekman, A. W. Poikola, and M. K. Mäkäräinen, "Invisible ENI: Using gaze and pupil size to control a game," in *Proc. CHI Extended Abstr. Human Factors Comput. Syst.*, 2008, pp. 3135–3140.

[34] I. Ekman, A. Poikola, M. Mäkäräinen, T. Takala, and P. Hämäläinen, "Voluntary pupil size change as control in eyes only interaction," in *Proc. Symp. Eye Tracking Res. Appl. (ETRA)*, 2008, pp. 115–118.

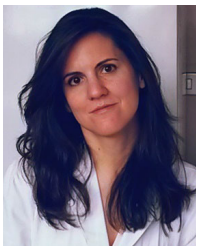
[35] A. E. Lorenzo Villalobos, S. Giusiano, L. Musso, C. de'Sperati, A. Riberi, P. Spalek, A. Calvo, C. Moglia, and S. Roatta, "When assistive eye tracking fails: Communicating with a brainstem-stroke patient through the pupillary accommodative response—A case study," *Biomed. Signal Process. Control*, vol. 67, May 2021, Art. no. 102515.



GIOVANNI CHIARION received the M.S. degree in biomedical engineering from the Polytechnic University of Turin, Italy, in 2019. He is currently pursuing the Ph.D. degree in electrical, electronics and communications engineering with the Polytechnic University of Turin. His research interests include functional connectivity analysis in neuroscience, involving signal and image processing techniques, and statistical data analysis.

FRANCESCO PONZIO (Member, IEEE) received the Ph.D. degree in control and computer engineering from the Polytechnic University of Turin, Italy, in 2020. He is currently a Postdoctoral Researcher with the Polytechnic University of Turin. His research interests include artificial intelligence, machine learning, and image processing in medical and smart manufacturing applications.

STEFANO TERNA received the M.S. degree in physics from the University of Torino, in 1996, and the Ph.D. degree in theoretical physics from the Scuola Internazionale Superiore di Studi Avanzati, Trieste, in 2000. From 2000 to 2004, he was a Senior Technology Consultant at the Accenture Financial Services Business Unit. From 2004 to 2008, he was the Software Product Manager at SAET IS, Torino, developing distributed systems for the detection and remote control of industrial alarms. From 2008 to 2016, he led, as a Freelancer, more than 30 software projects about field data acquisition, data analysis and remote control in automotive, banking, pharma, energy, and telecommunication sectors. Since 2015, he has been a CTO and a Co-founder with TomorrowData, a company specialized in developing and deploying AI algorithms on constrained hardware (the so-called edge computing). His research interests include recurrent neural networks, reservoir computing applied to the forecasting, and anomaly detection of multivariate time series. He was a member of the Advisory Board for the Master of Science in data analytics at the University of Central Florida, in 2016.



CRISTINA MOGLIA received the M.D. and Ph.D. degrees. She is currently an Assistant Professor of neurology with the Department of Neuroscience, University of Turin. She has a wide experience in the clinical management and research at ALS, since, she has been a part from over 15 years of ALS Expert Centre, Turin (CRESLA), AOU Città della Salute e della Scienza, Torino, which is the referral center for ALS, Piemonte Region, (Head: Prof Adriano Chiò). Her research interests include clinical, neuropsychological, genetic and epidemiological projects on ALS and other motor neuron disorders.

EDOARDO PATTI (Member, IEEE) received the M.Sc. and Ph.D. degrees in computer engineering from the Politecnico di Torino, Italy, in 2010 and 2014, respectively. He is currently an Associate Professor with the Politecnico di Torino. His research interests include ubiquitous computing, the Internet of Things, and smart systems applications.

SANTA DI CATALDO (Member, IEEE) received the Ph.D. degree in computer and systems engineering from Politecnico di Torino, in 2011. She is currently an Associate Professor with the Department of Control and Computer Engineering, Politecnico di Torino, Italy. Her research interests include artificial intelligence, machine learning, and image processing, with special regards to medical and smart manufacturing applications.

SILVESTRO ROATTA received the degree in electrical engineering from Politecnico di Torino, in 1993, and the Ph.D. degree in physiology from the University of Torino, in 1997. He is currently an Associate Professor of physiology and the Head of the Integrative Physiology Laboratory, with the Department of Neuroscience, University of Torino. He teaches human physiology at the School of Dentistry, Biomedical Engineering. He published over 70 articles on peer-reviewed journals. His research interests include motor control, hemodynamic measurements, muscle blood flow and performance, and human-computer interfaces.

Open Access funding provided by 'Politecnico di Torino' within the CRUI CARE Agreement