sensors

MDPI

Review

Augmentative and Alternative Communication (AAC) Advances: A Review of Configurations for Individuals with a Speech Disability

Yasmin Elsahar, Sijung Hu *, Kaddour Bouazza-Marouf, David Kerr and Annysa Mansor

Wolfson School of Mechanical, Electrical, and Manufacturing Engineering, Loughborough University, Loughborough LE11 3TU, UK; Y.Elsahar@lboro.ac.uk (Y.E.); K.Bouazza-marouf@lboro.ac.uk (K.B.-M.); D.Kerr@lboro.ac.uk (D.K.); annysa.mansor@gmail.com (A.M.)

* Correspondence: S.Hu@lboro.ac.uk; Tel.: +44-150-922-7058

Received: 13 March 2019; Accepted: 18 April 2019; Published: 22 April 2019

check for updates

Abstract: High-tech augmentative and alternative communication (AAC) methods are on a constant rise; however, the interaction between the user and the assistive technology is still challenged for an optimal user experience centered around the desired activity. This review presents a range of signal sensing and acquisition methods utilized in conjunction with the existing high-tech AAC platforms for individuals with a speech disability, including imaging methods, touch-enabled systems, mechanical and electro-mechanical access, breath-activated methods, and brain–computer interfaces (BCI). The listed AAC sensing modalities are compared in terms of ease of access, affordability, complexity, portability, and typical conversational speeds. A revelation of the associated AAC signal processing, encoding, and retrieval highlights the roles of machine learning (ML) and deep learning (DL) in the development of intelligent AAC solutions. The demands and the affordability of most systems hinder the scale of usage of high-tech AAC. Further research is indeed needed for the development of intelligent AAC applications reducing the associated costs and enhancing the portability of the solutions for a real user's environment. The consolidation of natural language processing with current solutions also needs to be further explored for the amelioration of the conversational speeds. The recommendations for prospective advances in coming high-tech AAC are addressed in terms of developments to support mobile health communicative applications.

Keywords: augmentative and alternative communication; assistive technologies; sensing modalities; signal processing; voice communication; machine learning; mobile health; speech disability

1. Introduction

Recent studies show that up to 1% of the world population suffers a degree of speech, language or communication need (SLCN) [1,2]. The loss of speech capabilities associated with extreme forms of paralysis and further medical complications has long been regarded as a barrier between the sufferers and the outside world. Augmentative and alternative communication (AAC) incorporates a wide range of processes that augment, complement, or replace speech of individuals with complex communication needs [3,4]. In the broad context of speech and language, *speech* is often associated with the motor movements responsible for the production of spoken words, whereas *language* is associated with the cognitive processing skills of communication.

AAC solutions are classified into three categories: no-tech, low-tech, and high-tech AAC [4]. No-tech AAC is considered the oldest of the three AAC categories, given its reliance on the interpretation of facial expressions and voluntary motor movements, such as sign language, to deliver non-verbal messages [5]. Low-tech AAC utilizes basic tools, such as books and display boards

3. Sensing Modalities and Their Functionalities

The integration of smart developments into daily life activities has widened the scope of dedicated and non-dedicated AAC applications [7,19]. A survey of high-tech AAC devices with regards to the signal acquisition, ML, and output generation is presented in this section.

3.1. AAC Signal Sources and Associated Processing

AAC interfaces are activated through an array of methods for the detection of human signals generated via body movements, respiration, phonation, or brain activities [4]. The acquisition of AAC signals is accomplished through several modalities. Table 1 outlines the AAC signal sensing categories discussed in this review together with their relevant activation methods. The listed AAC access methods could be used in a stand-alone format or in combination with one another. For example, imaging methods may be combined with touch-activated methods or mechanical switches to provide the users with a multi-modal access using the same device. A commercial example is Tobii Dynavox PCEye Plus, which combines several functionalities including eye tracking and switch access to use a computer screen [20].

Table 1. Sensing modalities of AAC signals.

Signal Sensing Category	Activation Method
Imaging methods	Eye gaze systems, head-pointing devices
Mechanical and Electromechanical methods	Mechanical keyboards, switch access
Touch-activated methods	Touchscreens, touch membrane keyboards
Breath-activated methods	Microphones, low-pressure sensors
Brain–Computer Interface methods	Invasive and non-invasive

3.1.1. Imaging Methods

Imaging methods, such as eye gazing, eye tracking and head-pointing devices, have been widely reported in the literature [21–31]. Eye gaze technologies work using the principle of tracking the eye movements of a user for the determination of the eye gaze direction [24,27]. Several eye tracking methods are commonly used, including video-oculography [32], electro-oculography [33], contact lenses [34], and electromagnetic scleral coils [21,25,30,35,36]. Oculography is involved with the measurement and recording of a user's eye movements [35]. Video-oculography and electro-oculography use video-based tracking systems and skin surface electrodes, respectively, to track the movements of the eye [25]. In the context of AAC, non-invasive eye tracking methods are better suited to address the daily needs of the users who lack motor abilities [27,29]. Practical methods involve the utilization of non-invasive cameras, an illumination source, image processing algorithms, and speech synthesizers to communicate a user's message [25,27]. Image data are obtained in video-oculography-operated systems using one or more cameras [23,27]. Typical video-oculography systems use glints produced on the surface of the eye through an illumination source, such as near-infrared (NIR) LEDs with typical wavelengths of 850 ± 30 nm, and in turn, gaze locations are estimated from the movement of the eye pupil in relation to the illuminated glint positions [34].

The components of a typical video-based tracking system are shown in Figure 2. Different approaches are presented in the literature for calculating the accuracy of an eye tracking system, including the distance accuracy (in cm or in pixels) and the angular accuracy (in degrees) [22]. The pixel accuracy can be given by

$$P_{acc} = \sqrt{\left(X_{target} PX\right)^2 + (Y_{target} PY)^2} \tag{1}$$

where X_{target} and Y_{target} are the coordinates of the target points, and PX and PY are the gaze point coordinates given by

$$PX = mean\left(\frac{PX_{left} + PX_{right}}{2}\right) \tag{2}$$

and

$$PY = mean\left(\frac{PY_{left} + PY_{right}}{2}\right) \tag{3}$$

respectively, with the subscripts $left$ and $right$ referring to the coordinates of gaze points of the left and right eyes. The on-screen distance accuracy (DA) is similarly given by

$$DA = p_{size}\sqrt{\left(PX - \frac{x_{pixels}}{2}\right)^2 + \left(y_{pixels} - PY + \frac{offset}{pixelsize}\right)^2} \tag{4}$$

where p_{size} is calculated based on the resolution, height, and width of the screen, x_{pixels} and y_{pixels} are the pixel shifts in the directions of x and y, respectively, and the $offset$ is the distance between the eye tracking unit and the lower edge of the screen [22,37]. The angular accuracy (AA) can be also computed via

$$AA = \frac{p_{size} \times P_{acc} \times cos(mean(\theta))^2}{meandist} \tag{5}$$

where the gaze angle θ is given by

$$\theta = tan^{-1}\left(\frac{DA}{dist}\right) \tag{6}$$

and $dist$ and $meandist$ are the distances from the eye to the screen and from the eye to the tracker, respectively [22,37].

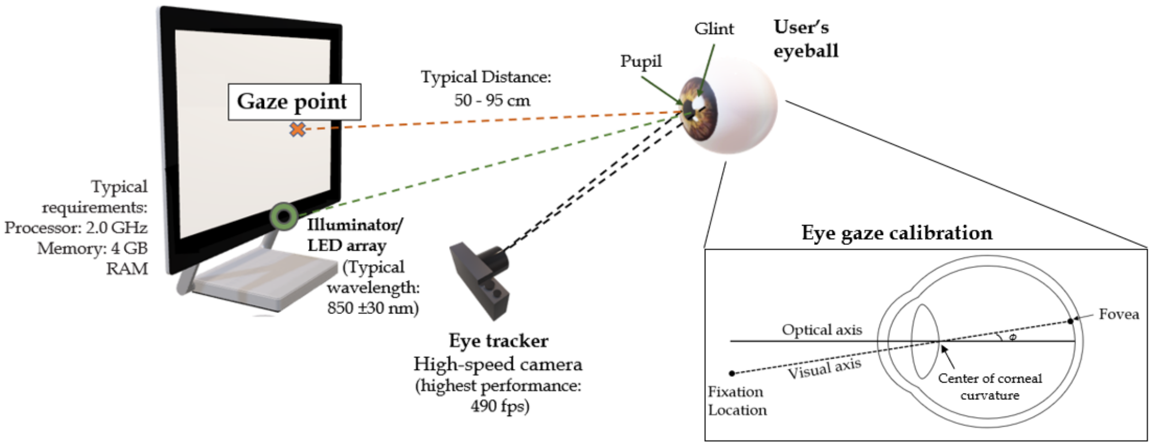

Figure 2. Components of a typical eye gaze system, adapted from [22,38]. The optical and the visual axes are used for the calibration process commonly required to set up the eye gaze system [22,39].

Fixations and saccades are commonly used to analyze eye movements [40]. Fixations are the pauses a user intently inputs by fixing his eye movements at the target gaze point, whereas saccades are the eye movements rapidly occurring following and in between the fixations. Metrics of eye gaze estimations include fixation durations, fixation rates, fixation sequences, saccadic amplitudes and velocities [22,40]. Although electro-oculography is a cost-effective eye tracking method, Infrared pupil corneal reflection (IR-PCR) video-based systems are most commonly used by speech and language practitioners due to their non-invasive nature [25,27]. A calibration operation is essential in video-based trackers to fine-tune the system with a user's eye movements [41]. As shown in Figure 2, a user's visual axis deviates from the optical axis upon the usage of a gaze system. Calibration is expressed as the process of finding the visual axis pertinent to each user by calculating the angle between the line joining the fovea (the highest point of sensitivity in the eye retina) with the center of corneal curvature, and the optical axis [22].