



Article

Improving the Reader's Attention and Focus through an AI-Driven Interactive and User-Aware Virtual Assistant for Handheld Devices

Giancarlo Iannizzotto ^{1,*}, Andrea Nucita ¹ and Lucia Lo Bello ²

¹ Department of Cognitive Sciences, Psychology, Education and Cultural Studies (COSPECS), University of Messina, 98122 Messina, Italy

² Department of Electrical, Electronic and Computer Engineering (DIEEI), University of Catania, 95124 Catania, Italy

* Correspondence: ianni@unime.it

Featured Application: This paper focuses on an AI-based application devised to be seamlessly integrated into tablets and e-book readers and aimed at helping users to improve their attention and focus on reading tasks.

Abstract: This paper describes the design and development of an AI-driven, interactive and user-aware virtual assistant aimed at helping users to focus their attention on reading or attending to other long-lasting visual tasks. The proposed approach uses computer vision and artificial intelligence to analyze the orientation of the head and the gaze of the user's eyes to estimate the level of attention during the task, as well as administer effective and balanced stimuli to correct significant deviations. The stimuli are provided by a graphical character (i.e., the virtual assistant), which is able to emulate face expressions, generate spoken messages and produce deictic visual cues to better involve the user and establish an effective, natural and enjoyable experience. The described virtual assistant is based on a modular architecture that can be scaled to support a wide range of applications, from virtual and blended collaborative spaces to mobile devices. In particular, this paper focuses on an application designed to integrate seamlessly into tablets and e-book readers to provide its services in mobility and exactly when and where needed.

Keywords: virtual intelligent assistant; artificial intelligence; handheld and mobile devices



Citation: Iannizzotto G.; Nucita A.; Lo Bello, L. Improving the Reader's Attention and Focus through an AI-Driven Interactive and User-Aware Virtual Assistant for Handheld Devices. *Appl. Syst. Innov.* **2022**, *5*, 92. <https://doi.org/10.3390/asi5050092>

Academic Editors: Wenbing Zhao and Andrey Chernov

Received: 11 July 2022

Accepted: 16 September 2022

Published: 22 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Computer-based aids demonstrated their effectiveness in helping users with individual study tasks in several cases [1–3]. Face tracking and other image analysis techniques were combined to understand students' reactions to different teaching methodologies in [4], eye gaze tracking technologies were recently used to assess student learning styles and personalize the content used for e-learning and remote lecturing [5], and the application of eye tracking to touchless exploration of graphical content was investigated in [6].

In recent years, high-end technologies such as augmented and virtual reality [7] and natural language recognition applied to voice assistants and smart speakers [8] moved from research laboratories to the realms of mobile and consumer electronics and industrial applications (Industry 4.0) [9], thus providing powerful foundations for innovative and, in some cases, disruptive approaches to learning and training. However, some negative effects of the massive use of digital media were also identified [10]. In particular, younger students still need to learn to focus their attention on "conventional" cognitive tasks such as reading and solving mathematical and geometrical problems. As a consequence, in addition to enabling innovative approaches to learning, computer-based technologies should support activities that aim to develop and improve the "conventional" cognitive abilities of humans at any age.

This article describes the “The Fun They Had” (TFTH) software platform (“The Fun They Had” is a short story written by Isaac Asimov [11] that describes a future world in which young students attend their individual school lectures at home and teachers are robots.) and a novel application named the Mobile Interactive Coaching Agent (MICA), aimed at helping the user improve his or her focus on a reading task. The application, built upon the TFTH platform, is especially designed for handheld devices such as tablets and e-book readers. The MICA exploits AI-based interaction technologies, such as face detection and recognition, head pose detection, eye gaze tracking, speech synthesis, and speech recognition, to estimate the degree of the user’s attention on reading and provide adequate gestural and verbal feedback. Furthermore, due to the underlying TFTH platform, the MICA supports other functions that are typical of widely adopted smart speakers, such as natural speech understanding and the ability to connect to smart home software and network services by integrating some functional modules derived from the Red platform, described in [12].

Although the Red and TFTH platforms share some properties and a modular architecture, they are very different systems. In fact, the Red platform was designed to support smart home applications, while TFTH has a different objective, that being mobile assistive and coaching applications and supporting the estimation of the cognitive state of the user, which is currently considered an important feature for future e-learning solutions [13]. Moreover, TFTH supports specific functionalities, such as head pose estimation and eye gaze estimation, which are not available on the Red platform.

The main contributions of this work are the following:

- We describe the modular software platform TFTH, which enables the visual observation of the user, the online analysis of the acquired data, the generation of different human-like gestural and spoken feedback, and the connection to other “smart assistant” services. The platform runs entirely on the device; that is, the processing needed to observe the user, analyze the data and produce the feedback is performed locally without accessing remote resources.
- We describe a coaching application, called the Mobile Interactive Coaching Agent (MICA), which leverages the services of the TFTH platform to monitor the user while reading text or attending to some other screen-based cognitive task and helps the user focus his or her attention on the task at hand.

2. Related Work

The idea of using graphical animated characters to attract and drive the user’s attention to specific items or areas on a reading surface (e.g., specific areas of a computer screen) has been investigated for decades. This derives directly from the observation that adding animated or salient patterns (such as the face of a virtual character or some attractive colored shape) to a region of an image is generally sufficient to attract the user’s attention to that region [14]. However, more recently, some research works investigated the ability of animated virtual characters to drive the user’s attention toward a specific direction by means of deictic cues (i.e., by pointing with gestures or with an eye gaze in the direction of the area of interest) [15]. Such an ability suggests that virtual characters may have a much more important role than generic moving graphical patterns. In fact, humans (as well as several primates and even some non-primates [16]) process deictic signals as nonverbal messages that are peculiar to living beings. As a consequence, the fact that the users tend to follow the deictic cues generated by animated graphical characters (often named virtual agents (VAs)) may also imply that users attribute a certain degree of liveliness and credibility to those characters. This hypothesis is intriguing because it opens up a wide range of possibilities worth exploring. In particular, the idea of emotional agents capable of recognizing the user’s face expressions and showing an adequate emotional response through synthetic face expressions and body gestures was investigated in [17,18].

Despite the wide interest generated in the research community and the promising results reported in the literature, after two decades, the actual diffusion of interactive

virtual agents (IVAs) is still limited, and the exploration of their potential is still far from complete [19]. In most cases, IVAs are implemented as graphical virtual characters that appear on the side of the main window of some application and are barely capable of measuring the time it takes for the user to read a page or complete a test. In contrast, thus far, only a few experiments have been conducted on fully functional agents capable of “sensing” users and reacting based on an estimate of the user’s state of mind, attention, interest in the activity at hand and fatigue. Exceptions to this lack of experimentation are confined mainly to areas of cognitive and sensorimotor rehabilitation, where some experimental studies have been performed [20]. Interestingly, AI-based technologies for sensing humans and responding to their cognitive and emotional states have been widely exploited in assistive and coaching robots for children affected by autism [21,22]. Robots powered by AI-based technologies for human and context sensing [23], as well as collaborative robots acting as team players [24], have been proven to be very effective in involving preschool children, for example, in play-like motor tasks.

The effectiveness of IVAs (and their robotic cousins) for specific but very diverse classes of users suggests that a reasonable degree of effectiveness should also characterize their application to more general subjects. In particular, an IVA capable of sensing the user could detect the most effective motivating factors to attract and keep their attention focused on the task at hand, thus dramatically improving their productivity [25,26]. Moreover, the recent widespread introduction of e-learning platforms and multimedia e-books for educational purposes, with the explicit objective of substituting, fully or partially, conventional textbooks, suggests that an experimentation of IVAs helping the user to focus their attention better and for a longer time on his or her tasks would produce significant and encouraging results. However, to the best of our knowledge, an open and technologically updated platform enabling experimentation in this field is currently not available.

IVAs need to sense the users (i.e., measure a set of physical and behavioral parameters in order to estimate some specific aspects of their state of mind toward the task at hand and (possibly) the environment), with the ultimate objective of improving users’ productivity, comfort and satisfaction or, in some cases, to allow them to interact with the environment when other means of interaction are lacking or insufficient [27]. To support such sensing abilities, current technologies exploit cameras, eye-gaze tracking devices, microphones, MEMS (3D acceleration sensors) and, in some cases, heart rate and blood oxygen saturation measurements from smartbands and smartwatches. Video data streams can be processed by AI algorithms to recognize the user (to differentiate between different users) and extract head poses and facial expressions [28] and, in some cases, track the eye gaze instead of using specialized eye-tracking hardware [29], recognize hand gestures and estimate and track full body poses and motions [30] when needed.

Current applications of AI-based computer vision, such as person detection, face detection and recognition, gesture recognition and pose detection, have been considered for years to be too demanding for handheld and wearable devices and were initially delegated to remote servers following a cloud-based paradigm. Recent improvements in the computational power of handheld and wearable devices and in the efficiency and reliability of AI algorithms allowed the execution of AI-based applications locally on a mobile device [12,31]. Local computation in turn reduces the amount of data transmitted from and to the mobile device, thus reducing power consumption and privacy-related risks. Moreover, local computation does not rely on wireless networks, which in some cases may be intrinsically unreliable.

In the context depicted, this work describes the architecture of a software platform named TFTH, which is aimed at developing AI-based applications capable of sensing and interacting with users, promoting their autonomy and productivity and running on mobile and handheld devices. Compared with the existing approaches in the literature, the innovative aspects proposed here are summarized as follows:

- The proposed solution operates locally on a handheld device and does not require remote (e.g., cloud-based) services for its core functions, such as face recognition, face detection, eye tracking, speech recognition and speech synthesis;
- The proposed solution does not require specialized hardware, such as eye gaze trackers, stereo cameras or dedicated processors (e.g., Intel Movidius) for data analysis.

The proposed platform has been used to develop the MICA pilot application, aimed at supporting users to focus their attention on a sustained visual task, such as reading. To assess the effectiveness of the MICA, in this work, we intend to answer the following research question:

- Does the MICA improve the user’s visual attention when reading from a handheld digital device?

We adopt a sustained visual attention model based on the assumption that the level of attention on a visual task is proportional to the percentage of time the user is looking at the object of the task. For example, when reading a text, the level of attention to the reading task is proportional to the percentage of time the user spends looking at the text being read.

3. Materials and Methods

The MICA was designed to run on a tablet computer device with average computational resources and a 10-inch color display. Nevertheless, its underlying TFTH modular platform, described in the following, also supports less powerful hardware platforms, such as some embedded systems or Raspberry III or Raspberry IV computers, and allows the development of more complex applications involving more functions and services.

3.1. The TFTH Architecture

For each basic function, in TFTH, several functional modules are available that differ in performance, memory footprint, computational load and other characteristics. The choice among the different modules available for each function is driven by design constraints.

Figure 1 shows the structure of the general TFTH architecture.

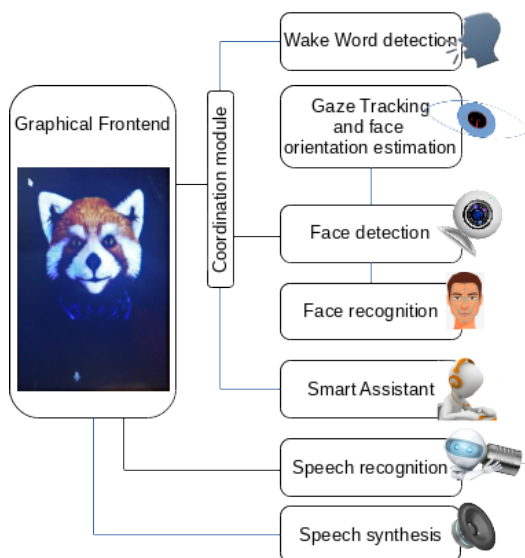


Figure 1. The TFTH architecture with the basic functional modules and one of the supported graphical characters.

The basic functional modules are devoted to speech input and output, face detection and recognition, estimation of head orientation and gaze tracking, natural language understanding and a smart assistant. The latter is responsible for connecting to smart home applications and devices and offers some Internet-based services similar to those supported by the most common smart speakers. A wake-word detection module is therefore added

to allow the user to initiate a conversation with the IVA whenever speech interaction is used. The graphical front-end is based on a 3D-animated character that shows facial expressions, makes head gestures and approximately moves its mouth according to the spoken sentences thanks to an ad hoc algorithm. The set of predefined facial expressions and head gestures is expandable, but those already available, derived from [12], are sufficient for most applications. The choice of character can arguably influence the effectiveness of the IVA [32], and therefore, several different characters were developed and are available for experimentation.

Finally, a coordination module is in charge of making decisions based on the data made available by the other modules. Depending on the application, the coordination module uses a suitable combination of input and output modules. For example, in the specific case of the MICA application considered here, the coordination module analyzes the gaze of the eye and the position of the head to determine the actions to take. Currently, the decision is made according to very simple rules described in Section 3.2. However, more accurate approaches based on deep learning are being investigated.

The whole TFTH architecture was developed under the assumption that only local computation is desirable when user-related data are processed; that is, all the processing needed to interact with the user is supposed to run locally on the handheld device, and no data should be transferred to external servers or cloud services. An exception is acceptable only if the user explicitly requests it, such as when a Google or Wikipedia search is commanded. This assumption is motivated by the privacy and reliability reasons that are explained in [33]. As a consequence, the low computational complexity was one of the main constraints in the development of the TFTH modules.

The speech synthesis module is based on a software library called pyttts3 [34], which in turn leverages the speech synthesis engine available on the device (Espeak for Linux or SAPI5 for Windows) or, for Apple devices, on a small Python wrapper for the native TTS engine [35].

The speech recognition module is based on the VOSK library [36], which provides adequate recognition functionalities for our purposes. The same library is used by the wake-word detection module.

The face detection and recognition module, derived from the model described in [12], is based on the OpenCV library [37] for face detection and the Dlib library [38] for face recognition. Both libraries are open source and run smoothly on the considered handheld devices.

Gaze tracking is a fundamental function for TFTH-based mobile applications. In particular, the ability to track the eye's gaze allows the MICA application to determine where the user is looking and therefore estimate the level of attention that the user is paying to the reading activity. However, very few handheld devices include dedicated hardware for eye tracking, and moreover, most devices directly supporting eye tracking are specific aids for impaired people. Some recent mobile devices (tablet PCs and smartphones) do have specialized hardware for face recognition and tracking that might be used for eye tracking, but such a feature is not yet available on the most common and more affordable mobile devices. External (USB) eye trackers available on the market usually require proprietary drivers that do not provide sufficient support for handheld devices. Similar issues affect most commercially available software solutions. For this reason, in [39], we developed a software solution for eye gaze tracking designed to work in both natural and near-infrared light. The proposed solution does not rely on special hardware or head-mounted devices, as in [40–42] or in some commercial solutions [43].

Gaze trackers are, in most cases, based on conventional IR pupil or corneal reflection [44] or more complex computer vision algorithms (depending on the available hardware) [45–47]. The approach adopted in TFTH was directly derived from [39] and optimized to best meet the needs of mobile device users. In particular, the screen size is between 5 and 11 inches, and the orientation is usually vertical (portrait). Furthermore, other specific issues had to be addressed, such as the low resolution of the input eye image, the non-

uniform illumination and the dependency of the eye gaze direction on both the head and camera orientations. Figure 2 provides a visualization of the intermediate results of the eye tracker for testing and debugging purposes. The 7-inch display of a Linux-based netbook is shown, where three separate windows show the orientation of the face and of the left and the right eyes. In addition, a small blue circle shows where the eye gaze of the user is actually pointing. In the figure, the eye images are horizontally flipped, so the user was looking approximately to the right of the screen.

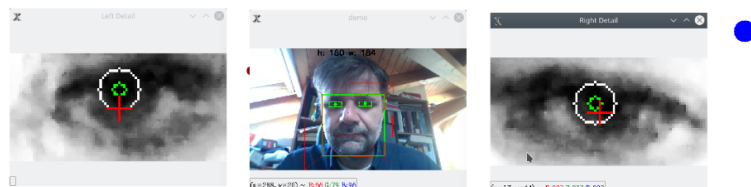


Figure 2. A screenshot demonstrating the adopted eye gaze tracking approach. The images are flipped horizontally. The blue circle marks the estimated target of the gaze on the screen.

The gaze tracker is intended to work on an average mobile CPU and does not require the support of a powerful GPU or specialized hardware for deep learning computation. To this end, it uses the face detector from the OpenCV library [37] and the face pose estimator and face parts detector from the Dlib library [38].

Figure 3 illustrates the very simple algorithm adopted. Once the pose of the head is estimated and the eyes are detected, each eye region undergoes histogram equalization and gray level filtering to improve the image quality. Then, a correlation-matching filter is applied which produces an image where the points that have a higher probability of being the centers of the eye pupil have higher brightness. This likelihood image is then weighted by a 2D Gaussian mask determined for each frame and each eye according to the size and shape of the eye image in that frame. Such a mask represents a “prior” that constrains the position of the pupil in the eye according to its size and shape. Finally, the point that features the maximum a posteriori (MAP) probability of being the center of the pupil is selected. The orientation angles of the two pupils are then averaged to obtain a more reliable estimate of the direction of the gaze of the eye. A five-fixed-point calibration procedure is applied to map the averaged gaze vectors of the two eyes on the device screen. The resolution obtained, by no means comparable to that of modern hardware-based gaze trackers, is sufficient for our purposes, as described in Section 3.2.

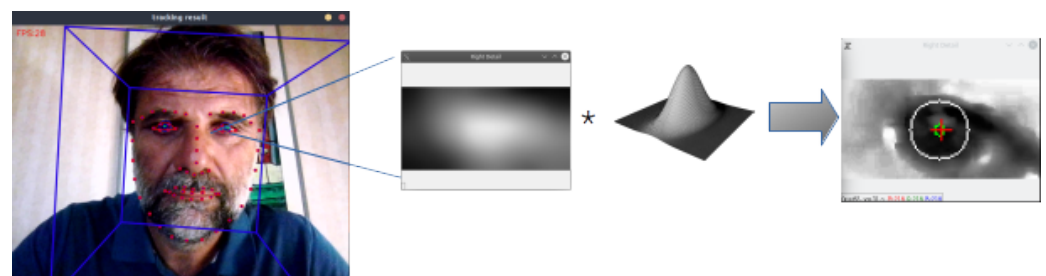


Figure 3. An illustration of the eye-tracking algorithm adopted (see text).

3.2. The Mobile Interactive Coaching Agent

The objective of the MICA is to help users improve their productivity and performance in reading and learning by better and more reliably focusing their attention. The underlying

rationale is that an intelligent and interactive graphical character (an IVA) capable of monitoring the user can be used as a pedagogical agent (PA) to increase learning motivation [32]. The role of the PA is supported by some cognitive and social theories that state that the user learns more deeply when the IVA on the screen displays human-like gesturing, movements, eye contact and facial expressions, in combination with just-in-time support or guidance to guide the user's attention to the key elements of the task [48–50].

For the purposes of this study, the MICA is content-agnostic and does not verify the actual level of understanding or memorization the user has of the content read. Further developments will consider the possibility of informing the MICA about the content that is administered to the user to improve the effectiveness of the coaching process.

Figure 4 shows the structure of the MICA image processing section. The face recognition module available in TFTH, based on the user recognition approach described in [12], identifies the user, thus allowing for per-user settings and calibration. This approach is common in recent operating systems such as Microsoft Windows 10 and is generally well accepted. However, it can be replaced, if needed, by more conventional user identification modules, such as $(user_id, password)$ credentials or fingerprint recognition.

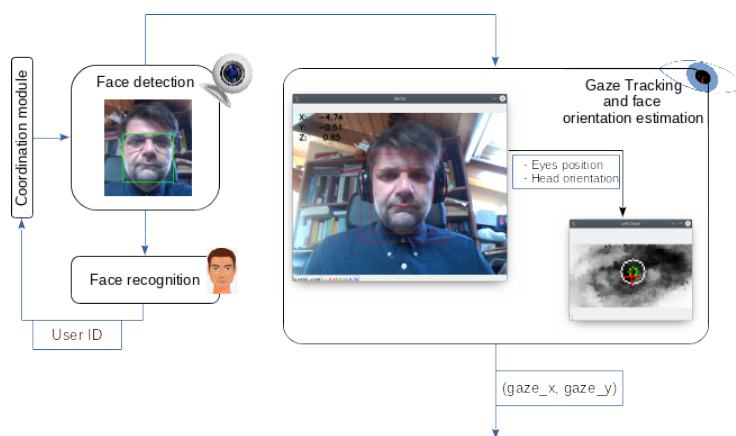


Figure 4. The MICA face processing architecture.

In [51], we showed how an interactive virtual agent capable of estimating the orientation of the user's face can infer a rough indication of their level of participation in the activity that is being carried out. Such an indication can, in turn, trigger a reaction of the agent, aiming at re-driving the user's focus on the activity at hand. The MICA combines the eye gaze with the face orientation to estimate the actual level of engagement and the focus of a user's attention. For example, while reading, the fixation point (i.e., the point in space that is being targeted by the combination of the two eyes) is supposed to be on the reading surface and is supposed to progress, on average, toward the end of the current page (not taking into account short, errant movements and saccades). Frequent and sensible deviations from this direction usually indicate some distraction or significant cognitive difficulties with the text in either understanding or remembering something. Instead, the comprehension of graphical content (figures, schemes, tables, etc.) usually requires different fixation patterns. For the sake of this research, the MICA adopts a content-agnostic mode. Therefore, the attention level is measured by detecting the *distraction events*, or the time intervals when the user's gaze focuses out of the screen area and estimating their durations. To avoid false alarms due to saccades and short, errant eye movements, only distraction events longer than a fixed heuristic threshold D_T , experimentally set to 0.9 s, were considered (i.e., approximately four times the average fixation time for silent reading) [52].

When the sum of the duration of all distraction events detected during a reading interval (for example, a reading interval of 60 s) exceeds a given threshold T_O , a *distraction warning* (DW) is raised. The MICA can manage a DW using two different approaches, according to the individual settings and (possibly) available information regarding the content being administered to the user. The first approach is to provide immediate feedback

to the user, asking them to refocus on the task. The adopted cumulative function of the frequency of distraction ensures that feedback is provided only after a significant amount of distraction has been detected, avoiding the risk that warnings that are too frequent may cause a distraction. The second approach is to provide feedback only at well-defined milestones for the activity, such as at the end of each chapter or lesson. This requires some information regarding the structure of the document being administered, which can be automatically extracted from the metadata when available. At each milestone, the user is provided with feedback on the attention paid to the content, and if a DW was raised, he or she receives the suggestion to revise some parts of the content, together with the specification of the parts needing revision. Figure 5 shows a very simple case of reading an e-book that takes advantage of the MICA to motivate the user and help focus their attention. The image on the left shows the MICA character while looking at the user. The expression of the character is neutral, although some occasional small movements of the head and eyes contribute to the “liveliness” of the character. On the right, the character uses speech synthesis and facial expressions to provide feedback to the user about his or her level of attention while reading.

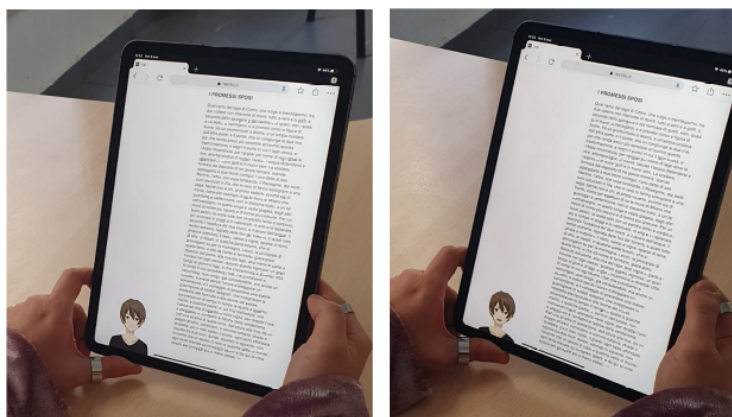


Figure 5. A simple reading application featuring MICA. **(Left)** The agent is observing the user. **(Right)** The agent is providing the user with a feedback about his or her level of attention while reading.

In specific cases, such as when the user is very young or has some cognitive impairments, it has been shown [51] that giving positive feedback (such as smiling and uttering encouraging sentences) to the user may improve his or her engagement and motivation. However, such feedback can be provided only at specific stages of the activity (e.g., at the end of a chapter or a lesson) when the metadata regarding the content are available. Moreover, there are cases (e.g., adult users) in which positive feedback should be avoided, as it may be a source of distraction or disturbance. As a consequence, although the MICA can provide positive feedback, this feedback is not emitted frequently.

The choice of the graphical character of the interactive agent was found to be critical for at least some categories of users [53,54]. As a consequence, the MICA has been designed to take on the appearance of different graphical characters according to the user’s preference, and a set of default characters was individuated following the preliminary investigations. Figure 6 shows some of the characters performing different facial expressions.



Figure 6. The three MICA tested characters.

In the MICA, the character animations are synchronized with the uttered speech, either synthetic or sampled, and show human-like facial expressions. Compared with previous work, deictic cues were added to the set of facial animations to facilitate interaction with the user during the coaching process, as shown in Figure 7.



Figure 7. MICA characters support deictic visual cues.

The first requirement for the MICA refers to its ability to effectively capture the dynamics of the user’s attentive state. In other words, the MICA needs to run fast enough to track the changes in the user’s attitude that might reveal significant changes in their attention. We chose the frame-processing rate in frames per second (f/s) to measure the processing speed. According to the literature (e.g., [55]), we decided that for the purpose of monitoring the user’s attention while reading, 10 FPS could be assumed as the lowest acceptable sampling speed.

To assess the ability of the presented approach to meet the first requirement, we ran the MICA on three different hardware architectures that represent a wide range of target devices, namely two different tablet PCs and a Raspberry Pi 3B+ board equipped with different ARM processors, different RAM capacities and different cameras. For the sake of comparability, we set the acquired video resolution to 1280 × 720 pixels. The main features of the devices tested and the frame rates obtained are reported in Table 1.

Table 1. Tested devices and obtained frame rates.

Device	CPU	RAM Capacity	Frame Rate
Tablet	A12X, 8 cores, 2.48 GHz	4 GB	24
Tablet	Kirin 659, 8 cores, 2.36 GHz	3 GB	15
Raspberry Pi 3B+	Cortex A53, 4 cores, 1.4 GHz	1 GB	10

Note that the current MICA implementation is not yet optimized for the different architectures. Therefore, its performance could improve with a more advanced prototype.

The second requirement is that the MICA must not sensibly impair the ability of the device to run the other applications that the user needs for the task at hand (e.g., e-book reading software or a web browser). We found that in all cases, neither the e-book reader application (the Amazon Kindle application for the two tablet PCs or the Calibre Ebook Viewer application for the Raspberry Pi 3B + board) nor the web browser were sensibly affected by the MICA in terms of reactivity or functionality.

3.3. Experimental Set-up

We set up a proof-of-concept experiment with real users to get a hint of the effectiveness of the application (i.e., its ability to help users focus their attention on the reading task).

A total of 10 male subjects and 10 female subjects were recruited from 6th to 8th grade in schools.

Taking into account the level of education of the subjects, six excerpts from the novel *I Promessi Sposi (The Betrothed)* by Alessandro Manzoni in its original Italian version were selected and named “E1”, “E2”, ... “E6”. Each excerpt was four pages long, text only and without figures.

We set up 20 tablets (one for each subject) with a web-based e-book reader application that supported the MICA (see Figure 4). Each subject was initially left with the device and the e-reader application running on a different text for 10 min to familiarize with the application and the presence of the MICA interactive character. A briefing was also provided, presenting the trial as a challenge and advising the subjects that they should not be distracted by the character and that staring at the character would reduce their performance.

Two different experimental conditions were devised. In the first condition (C1), a non-interactive animated character was shown on the screen that was graphically identical to the MICA character but unable to provide any feedback to the user. In the second condition (C2), the MICA character was shown and gave immediate feedback to the user. Immediate feedback, according to the definition in Section 3.2, was provided only after a given cumulative amount of distraction was detected.

After the initial familiarization interval, all subjects were asked to read the six excerpts in sequence in condition C1. For this task, 13 min per excerpt (approximately 3 min per page) were given plus a 1-min stop between two subsequent excerpts. Once all subjects completed their tasks, the first reading session was closed, and a second session was planned after a 1-day pause.

In the second reading session, all subjects were asked to read the six excerpts in sequence again but under condition C2. Again, 13 min per excerpt were given (approximately 3 min per page) plus a one-minute stop between two subsequent excerpts. Once all subjects completed their tasks, the second reading session was also closed, which also concluded the data gathering process.

For each trial (that is, a user reading an excerpt), the following set of measures was collected:

- The total reading time for the entire excerpt;
- The number of distraction events detected during the trial;
- The total reading time during which the user was distracted (i.e., the cumulative distraction time during the reading session).

In this proof-of-concept experiment, we did not take into account the structure of the document. As a consequence, a distraction was detected only when the user's eye gaze moved away from the page being read, including when the fixation moved from the text page to the animated graphical character, which resided on a different screen area. Other types of distraction symptoms, such as an eye's gaze wandering on the page or moving back and forth or staring at a specific area for a long time, were not considered.

4. Results

For the sake of clarity in data visualization, for each subject, the reading time, the number of distraction events and the cumulative distraction time were averaged over the six excerpts separately for the two conditions (C1 and C2), thus producing an average reading time, an average number of distraction events and an average cumulative distraction time for each subject and for each condition (C1 and C2). As a result, Figure 8 on the left compares the graphs of the average reading time for each subject in condition C1 (non-interactive graphical character) and condition C2 (MICA), while Figure 8 on the right compares the graphs of the average cumulative distraction time under the same two conditions.

The graphs show that substantial improvement was consistently obtained for both the overall reading time and cumulative distraction time for all subjects with the introduction of the MICA (i.e., in condition C2).

To better understand this improvement, the percentage reduction in reading time RTR_i in condition C2 was also calculated relative to condition C1 for each subject according to Equation (1), where $RT_i(C1)$ and $RT_i(C2)$ represent the average reading time for the subject i under conditions C1 and C2, respectively:

$$RTR_i = (RT_i(C1) - RT_i(C2)) / RT_i(C1) * 100, \forall i \in [1..20] \quad (1)$$

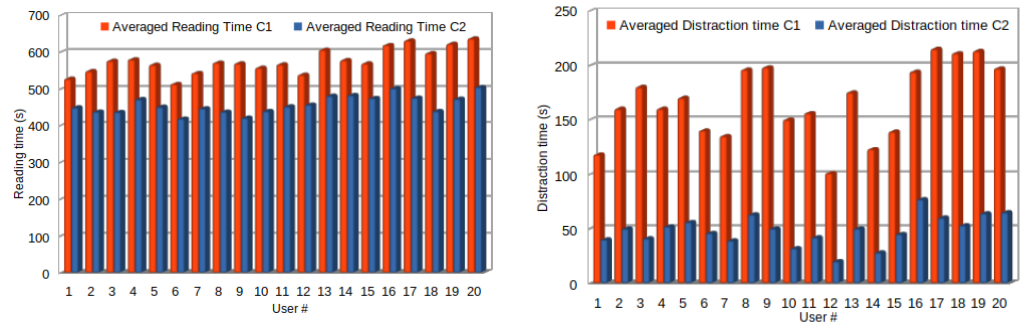


Figure 8. Plots of the reading time (on the left) and the cumulative distraction time (on the right) for each subject, averaged over the six excerpts in conditions C1 and C2 (see text).

Similarly, for each subject, the percentage reduction in distraction time DTR_i in condition C2 was calculated with respect to condition C1 according to Equation (2), where $DT_i(C1)$ and $DT_i(C2)$ represent the average cumulative distraction time for the subject i under conditions C1 and C2, respectively:

$$DTR_i = (DT_i(C1) - DT_i(C2)) / DT_i(C1) * 100, \forall i \in [1..20] \tag{2}$$

Figure 9 shows the graph of RTR_i (left) and the graph of DTR_i (right) for all subjects. The graph on the left clearly shows a significant improvement in the average reading time for all subjects, although with sensible individual variations. This improvement is largely explained by the graph on the right, which shows that the cumulative distraction time decreased by more than 60% and up to 80% for some subjects with the introduction of the MICA.

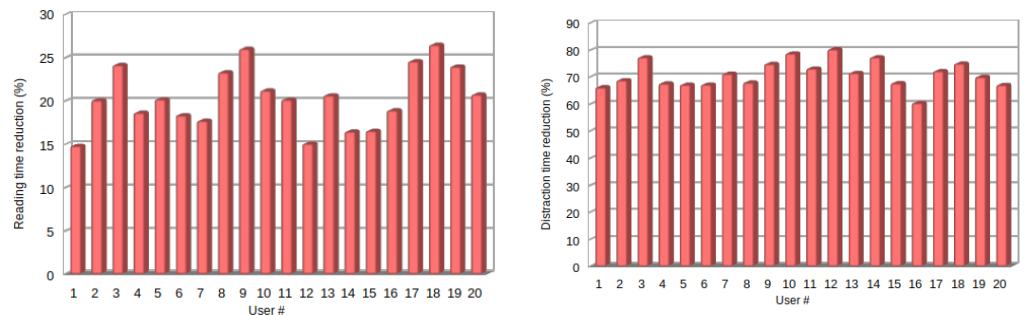


Figure 9. Plots of the percentage reduction in reading time (on the left) and percentage reduction in cumulative distraction time (on the right) for each subject, averaged over the six excerpts (see text).

To collect some subjective information on the user experience and the level of reliability of the experimentation, after the end of the two reading sessions, a very simple questionnaire was administered to the subjects. The questionnaire posed only one question, reported in Table 2.

Table 2. Question asked to the users.

While reading, the MICA made you more: Distracted Focused Indifferent

The answers obtained are reported in Table 3. Only 2 out of 20 subjects responded that the MICA left them indifferent, while no one declared that they had been distracted. In general, 90% of the subjects were satisfied and declared that the effect of the MICA was positive.

Table 3. Answers to the user experience questionnaire.

Distracted	Focused	Indifferent
0%	90%	10%

5. Discussion

The MICA encouraged the user to focus on the reading task, providing meaningful feedback through a voice, facial expression and head gestures. The feedback was based on the observation of the user during the working session through video analysis and by tracking their interactions with the device interface (touch screen). Through video analysis, the face of the user was constantly observed, and the orientation of the face and gaze of the eyes were estimated. Touch screen data were used to track the reading pace (i.e., how many pages were spanned in the last observation interval). The acquired data were then used to provide adequate and timely feedback to the user.

The feedback could be provided according to two different paradigms: either every time a significant loss of attention was detected or at predefined milestones, such as at the end of a lesson or a page. The two paradigms produced sensibly different effects. The timely feedback prevented the user from wasting further time due to a distraction, but it may have become annoying when either distractions were detected too frequently or some behavior of the user was wrongfully detected as a loss of attention. On the other hand, cumulative feedback can be provided to make the user more responsible, such as by asking them some questions that assess the level of content comprehension achieved. In most cases, it is not necessary to analyze the provided answer, as the effort needed to answer the question will be sufficient to convince the user that some content review is needed. Currently, no definitive response from the experimental data has emerged as to which paradigm is better for different situations, and therefore, further investigations will be carried out.

The data acquired while monitoring the user can also be exploited to assess the improvement produced by some treatment administered through the device on which the agent runs. For example, if the device is used to provide some content for cognitive rehabilitation, then the MICA can be used to motivate the user and, at the same time, collect observational data useful for evaluating the improvements introduced by the treatment or, preliminarily, to detect some cognitive problems [56]. Furthermore, the MICA is mobile-oriented, and therefore its architecture adheres to how people are most likely to consume books and other multimedia content (i.e., using mobile devices) [57].

We organized a simple but effective experimental campaign to evaluate the proposed approach, performing 240 reading tests with 20 different subjects, totaling approximately 52 h of reading time. The reading sessions were designed to produce a sensible level of fatigue in the subjects, since each subject was asked to read 6 excerpts in sequence (with a 1-min stop between consecutive readings) for a total of approximately 78 min for each reading session. The results presented clearly show that the MICA greatly improved the visual attention of the subjects during the reading task.

6. Limitations and Future Research

The described approach has some limitations that must be taken into account. First, a very simple model of attention was adopted, considering the eye's gaze leaving the reading area as the only symptom of distraction. As stated above, the readers might be distracted even though their eye gazes were still fixated on the reading area, such as wandering randomly on it or moving backward and forward. If accurate attention level measurement is required, more effective and reliable measures are needed, such as one based on an evaluation of the understanding of the text read. To better monitor the real visual attention of the reader, a more effective approach, currently under investigation as a future improvement to the MICA, is to extract some metadata regarding the structure of the document, such as the division into paragraphs, the structure of each page and the presence

of graphical elements on the current page to estimate an expected visual path in the current page and to evaluate the consistency of the tracked eye gaze path with the estimated visual path, which is roughly similar to some visual trajectory analysis methods [58].

A second limitation of the MICA is the low resolution of its eye gaze tracker, which is currently barely capable of discriminating between inside and outside the reading area. However, recent deep convolutional neural networks (CNNs) have been shown to be very effective in face and head detection, face tracking and eye detection [45,59]. As a consequence, more accurate eye tracking approaches will be investigated in the near future, taking advantage of the most recent lightweight implementations of deep learning platforms such as TensorFlow Lite [60], Pytorch [61] and OpenVino [62].

Finally, the MICA is defined as an “interactive” agent because it actually observes the user, estimates his or her cognitive state and provides “human-like” feedback in a continuous loop throughout the reading task. More articulated interactions are, in fact, supported by the TFTH platform, which involves the interpretation and synthesis of natural spoken language, the interpretation and synthesis of facial expressions and other abilities inherited from the Red platform [12]. However, it has not yet been demonstrated that introducing all those communication channels actually improves user attention, and therefore we plan to gradually test them in the future.

7. Conclusions

This paper introduced the software architecture of an interactive virtual agent that can support the user who is reading, attending an e-learning lesson or receiving cognitive rehabilitation treatment. The agent, named the Multimodal Interactive Coaching Agent (MICA), senses the user by visually observing his or her face, head orientation and eye gaze and estimates the level of attention and motivation. In response to the estimated state, the agent provides the user with specific multimodal stimuli using speech, facial expressions, head gestures and deictic visual cues.

The MICA has been tested with 20 different users and more than 50 h of reading sessions, producing a significant amount of data. Analysis of the experimental results shows that the MICA introduced an important improvement in the visual attention of the user toward the reading task, thus proving the validity of the proposed approach.

Future research directions are proposed with regard to the application of recent deep learning computer vision approaches to eye gaze tracking and the adoption of a more detailed and reliable visual attention model.

Author Contributions: Conceptualization, G.I.; methodology, G.I.; validation, L.L.B.; formal analysis, G.I.; investigation, G.I.; data curation, A.N.; writing—original draft preparation, G.I.; writing—review and editing, A.N. and L.L.B.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cinquin, P.; Guitton, P.; Sauz on, H. Online e-learning and cognitive disabilities: A systematic review. *Comput. Educ.* **2019**, *130*, 152–167. [[CrossRef](#)]
2. Stan in, K.; Hoi -Bo i , N.; Sko i  Mihi , S. Using digital game-based learning for students with intellectual disabilities—A systematic literature review. *Inform. Educ.* **2020**, *19*, 323–341. [[CrossRef](#)]
3. Coletta, A.; De Marsico, M.; Panizzi, E.; Prenkaj, B.; Silvestri, D. MIMOSE: Multimodal interaction for music orchestration sheet editors. *Multimed. Tools Appl.* **2019**, *78*, 33041–33068. [[CrossRef](#)]
4. Marceddu, A.C.; Pugliese, L.; Sini, J.; Espinosa, G.R.; Amel Solouki, M.; Chiavassa, P.; Giusto, E.; Montrucchio, B.; Violante, M.; De Pace, F. A Novel Redundant Validation IoT System for Affective Learning Based on Facial Expressions and Biological Signals. *Sensors* **2022**, *22*, 2773. doi: 10.3390/s22072773. [[CrossRef](#)]

5. Nugrahaningsih, N.; Porta, M.; Klasnja-Milicevic, A. Assessing learning styles through eye tracking for e-learning applications. *Comput. Sci. Inf. Syst.* **2021**, *18*, 1287–1309. [[CrossRef](#)]
6. Dondi, P.; Porta, M.; Donvito, A.; Volpe, G. A gaze-based interactive system to explore artwork imagery. *J. Multimodal User Interfaces* **2022**, *16*, 55–67. [[CrossRef](#)]
7. Batista, A.F.; Thiry, M.; Gonçalves, R.Q.; Fernandes, A. Using technologies as virtual environments for computer teaching: A systematic review. *Inform. Educ.* **2020**, *19*, 201–221. [[CrossRef](#)]
8. Terzopoulos, G.; Satratzemi, M. Voice assistants and smart speakers in everyday life and in education. *Inform. Educ.* **2020**, *19*, 473–490. [[CrossRef](#)]
9. Okwu, M.O.; Tartibu, L.K.; Maware, C.; Enarevba, D.R.; Afenogho, J.O.; Essien, A. Emerging Technologies of Industry 4.0: Challenges and Opportunities. In Proceedings of the 2022 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), Durban, South Africa, 4–5 August 2022; pp. 1–13. doi: 10.1109/icABCD54961.2022.9856002. [[CrossRef](#)]
10. Capri, T.; Gugliandolo, M.C.; Iannizzotto, G.; Nucita, A.; Fabio, R. The influence of media usage on family functioning. *Curr. Psychol.* **2021**, *40*, 2644–2653. doi: 10.1007/S12144-019-00204-1. [[CrossRef](#)]
11. Asimov, I. The Fun They Had. *Mag. Fantasy Sci. Fict.* **1954**, *6*, 125.
12. Iannizzotto, G.; Lo Bello, L.; Nucita, A.; Grasso, G.M. A Vision and Speech Enabled, Customizable, Virtual Assistant for Smart Environments. In Proceedings of the 2018 11th International Conference on Human System Interaction (HSI), Gdansk, Poland, 4–6 July 2018; pp. 50–56. doi: 10.1109/HSI.2018.8431232. [[CrossRef](#)]
13. Ivanova, M. eLearning informatics: From automation of educational activities to intelligent solutions building. *Inform. Educ.* **2020**, *19*, 257–282. [[CrossRef](#)]
14. De Koning, B.B.; Tabbers, H.K.; Rikers, R.M.J.P.; Paas, F. Towards a Framework for Attention Cueing in Instructional Animations: Guidelines for Research and Design. *Educ. Psychol. Rev.* **2009**, *21*, 113–140. [[CrossRef](#)]
15. Martinez, S.; Sloan, R.; Szymkowiak, A.; Scott-Brown, K. Animated virtual agents to cue user attention: Comparison of static and dynamic deictic cues on gaze and touch responses. *Int. J. Adv. Intell. Syst.* **2011**, *4*, 299–308.
16. Shepherd, S.V. Following gaze: Gaze-following behavior as a window into social cognition. *Front. Integr. Neurosci.* **2010**, *4*, 5. [[CrossRef](#)]
17. Morishima, Y.; Nakajima, H.; Brave, S.; Yamada, R.; Maldonado, H.; Nass, C.; Kawaji, S. The Role of Affect and Sociality in the Agent-Based Collaborative Learning System. In *Affective Dialogue Systems Proceedings of the Tutorial and Research Workshop, ADS 2004, Kloster Irsee, Germany, 14–16 June 2004*; André, E., Dybkjær, L., Minker, W., Heisterkamp, P., Eds.; Springer: Berlin/Heidelberg, Germany, 2004; pp. 265–275.
18. Ammar, M.; Neji, M.; Alimi, A.M. The Role of Affect in an Agent-Based Collaborative E-Learning System Used for Engineering Education. In *Handbook of Research on Socio-Technical Design and Social Networking Systems*; Whitworth, B., de Moor, A., Eds.; IGI Global: Hershey, PA, USA, 2009.
19. Johnson, W.L.; Lester, J.C. Face-to-Face Interaction with Pedagogical Agents, Twenty Years Later. *Int. J. Artif. Intell. Educ.* **2016**, *26*, 25–36. doi: 10.1007/s40593-015-0065-9. [[CrossRef](#)]
20. Bernardini, S.; Porayska-Pomsta, K.; Smith, T.J. ECHOES: An intelligent serious game for fostering social communication in children with autism. *Inf. Sci.* **2014**, *264*, 41–60. doi: 10.1016/j.ins.2013.10.027. [[CrossRef](#)]
21. Clabaugh, C.; Mahajan, K.; Jain, S.; Pakkar, R.; Becerra, D.; Shi, Z.; Deng, E.; Lee, R.; Ragusa, G.; Matarić, M. Long-Term Personalization of an In-Home Socially Assistive Robot for Children With Autism Spectrum Disorders. *Front. Robot. AI* **2019**, *6*, 110. doi: 10.3389/frobt.2019.00110. [[CrossRef](#)] [[PubMed](#)]
22. Rudovic, O.; Lee, J.; Dai, M.; Schuller, B.; Picard, R.W. Personalized machine learning for robot perception of affect and engagement in autism therapy. *Sci. Robot.* **2018**, *3*, eaao6760. doi: 10.1126/scirobotics.aa6760. [[CrossRef](#)]
23. Fridin, M.; Belokopytov, M. Embodied Robot versus Virtual Agent: Involvement of Preschool Children in Motor Task Performance. *Int. J. Hum.-Comput. Interact.* **2014**, *30*, 459–469. doi: 10.1080/10447318.2014.888500. [[CrossRef](#)]
24. Patti, G.; Leonardi, L.; Lo Bello, L. A Novel MAC Protocol for Low Datarate Cooperative Mobile Robot Teams. *Electronics* **2020**, *9*, 235. doi: 10.3390/electronics9020235. [[CrossRef](#)]
25. Qu, L.; Wang, N.; Johnson, W.L. Using Learner Focus of Attention to Detect Learner Motivation Factors. In *User Modeling 2005 Proceedings of the 10th International Conference, UIM 2005, Edinburgh, Scotland, UK, 24–29 July 2005*; Ardissono, L., Brna, P., Mitrovic, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; pp. 70–73.
26. Frémont, V.; Phan, M.T.; Thouvenin, I. Adaptive Visual Assistance System for Enhancing the Driver Awareness of Pedestrians. *Int. J. Hum.-Comput. Interact.* **2020**, *36*, 856–869. doi: 10.1080/10447318.2019.1698220. [[CrossRef](#)]
27. Fabio, R.; Capri, T.; Nucita, A.; Iannizzotto, G.; Mohammadhasani, N. Eye-gaze digital games improve motivational and attentional abilities in RETT syndrome. *J. Spec. Educ. Rehabil.* **2019**, *19*, 105–126. [[CrossRef](#)]
28. Kazemi, V.; Sullivan, J. One millisecond face alignment with an ensemble of regression trees. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1867–1874. doi: 10.1109/CVPR.2014.241. [[CrossRef](#)]
29. Papoutsaki, A.; Sangkloy, P.; Laskey, J.; Daskalova, N.; Huang, J.; Hays, J. Webgazer: Scalable Webcam Eye Tracking Using User Interactions. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, New York, NY, USA, 9–15 July 2016; pp. 3839–3845.

30. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.E.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 172–186. doi: 10.1109/TPAMI.2019.2929257. [CrossRef] [PubMed]
31. Papandreou, G.; Zhu, T.; Chen, L.C.; Gidaris, S.; Tompson, J.; Murphy, K. PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model. In *Proceedings of the Computer Vision—ECCV 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 282–299.
32. Mohammadhasani, N.; Fardanesh, H.; Hatami, J.; Mozayani, N.; Fabio, R.A. The pedagogical agent enhances mathematics learning in ADHD students. *Educ. Inf. Technol.* **2018**, *23*, 2299–2308. doi: 10.1007/s10639-018-9710-x. [CrossRef]
33. Iannizzotto, G.; Nucita, A.; Fabio, R.A.; Capri, T.; Lo Bello, L. More Intelligence and Less Clouds in Our Smart Homes. In *Economic and Policy Implications of Artificial Intelligence*; Marino, D., Monaca, M.A., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 123–136. doi: 10.1007/978-3-030-45340-4_9. [CrossRef]
34. Natesh Bhat. pyttax3 Library. Available online: <https://github.com/nateshmbhat/pyttax3/> (accessed on 11 September 2022).
35. yumoqing. ios_tts. Available online: https://github.com/yumoqing/ios_tts (accessed on 11 September 2022).
36. Alpha Cephei. pyttax3 Library. Available online: <https://github.com/alphacep/vosk-api> (accessed on 11 September 2022).
37. Itseez. Open Source Computer Vision Library. Available online: <https://github.com/itseez/opencv>. (accessed on 11 September 2022).
38. King, D.E. Dlib-ml: A Machine Learning Toolkit. *J. Mach. Learn. Res.* **2009**, *10*, 1755–1758.
39. Iannizzotto, G.; Nucita, A.; Fabio, R.A.; Capri, T.; Bello, L.L. Remote Eye-Tracking for Cognitive Telerehabilitation and Interactive School Tasks in Times of COVID-19. *Information* **2020**, *11*, 296. [CrossRef]
40. Wolfe, B.; Eichmann, D. A Neural Network Approach to Tracking Eye Position. *Int. J. Hum.-Comput. Interact.* **1997**, *9*, 59–79. doi: 10.1207/s15327590ijhc0901_4. [CrossRef]
41. Li, D.; Babcock, J.; Parkhurst, D.J. OpenEyes: A Low-Cost Head-Mounted Eye-Tracking Solution. In *Proceedings of the 2006 Symposium on Eye Tracking Research & Applications*, San Diego, CA, USA, 27–29 March 2006; Association for Computing Machinery: New York, NY, USA, 2006; ETRA '06, pp. 95–100. doi: 10.1145/1117309.1117350. [CrossRef]
42. Lee, K.F.; Chen, Y.L.; Yu, C.W.; Chin, K.Y.; Wu, C.H. Gaze Tracking and Point Estimation Using Low-Cost Head-Mounted Devices. *Sensors* **2020**, *20*, 1917. doi: 10.3390/s20071917. [CrossRef]
43. Kassner, M.; Patera, W.; Bulling, A. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-Based Interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, Seattle, WA, USA, 13–17 September 2014; Association for Computing Machinery: New York, NY, USA, 2014; UbiComp '14 Adjunct, pp. 1151–1160. doi: 10.1145/2638728.2641695. [CrossRef]
44. Duchowski, A.T. *Eye Tracking Methodology: Theory and Practice*; Springer: London, UK, 2003.
45. Valliappan, N.; Dai, N.; Steinberg, E.; He, J.; Rogers, K.; Ramachandran, V.; Xu, P.; Shojaeizadeh, M.; Guo, L.; Kohlhoff, K.; et al. Accelerating eye movement research via accurate and affordable smartphone eye tracking. *Nat. Commun.* **2020**, *11*, 4553. [CrossRef]
46. Baltrusaitis, T.; Zadeh, A.; Lim, Y.C.; Morency, L.P. OpenFace 2.0: Facial Behavior Analysis Toolkit. In *Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi'an, China, 15–19 May 2018; pp. 59–66. doi: 10.1109/FG.2018.00019. [CrossRef]
47. Iannizzotto, G.; La Rosa, F. Competitive Combination of Multiple Eye Detection and Tracking Techniques. *IEEE Trans. Ind. Electron.* **2011**, *58*, 3151–3159. doi: 10.1109/TIE.2010.2102314. [CrossRef]
48. Nass, C.; Brave, S. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*; MIT Press: Cambridge, MA, USA, 2007.
49. Bandura, A. Social Cognitive Theory of Mass Communication. *Media Psychol.* **2001**, *3*, 265–299. doi: 10.1207/S1532785 XMEP0303_03. [CrossRef]
50. Sweller, J. Cognitive load theory. In *Psychology of Learning and Motivation*; Academic Press: Cambridge, MA, USA, 2011; Volume 55, pp. 37–76. doi: 10.1016/B978-0-12-387691-1.00002-8. [CrossRef]
51. Fabio, R.A.; Capri, T.; Iannizzotto, G.; Nucita, A.; Mohammadhasani, N. Interactive Avatar Boosts the Performances of Children with Attention Deficit Hyperactivity Disorder in Dynamic Measures of Intelligence. *Cyberpsychol. Behav. Soc. Netw.* **2019**, *22*, 588–596. [CrossRef] [PubMed]
52. Rayner, K. Eye movements in reading and information processing. *Psychol. Bull.* **1978**, *85*, 618–660. [CrossRef]
53. Domagk, S. Do pedagogical agents facilitate learner motivation and learning outcomes? The role of the appeal of agent's appearance and voice. *J. Media Psychol. Theor. Methods Appl.* **2010**, *22*, 84. [CrossRef]
54. Veletsianos, G. Contextually relevant pedagogical agents: Visual appearance, stereotypes, and first impressions and their impact on learning. *Comput. Educ.* **2010**, *55*, 576–585. doi: 10.1016/j.compedu.2010.02.019. [CrossRef]
55. Landau, A.; Fries, P. Attention Samples Stimuli Rhythmically. *Curr. Biol.* **2012**, *22*, 1000–1004. doi: 10.1016/j.cub.2012.03.054. [CrossRef]
56. Mohammadhasani, N.; Capri, T.; Nucita, A.; Iannizzotto, G.; Fabio, R.A. Atypical Visual Scan Path Affects Remembering in ADHD. *J. Int. Neuropsychol. Soc.* **2020**, *26*, 557–566. [CrossRef]
57. Fabio, R.A.; Iannizzotto, G.; Nucita, A.; Capri, T. Adult listening behaviour, music preferences and emotions in the mobile context. Does mobile context affect elicited emotions? *Cogent Eng.* **2019**, *6*, 1597666. doi: 10.1080/23311916.2019.1597666. [CrossRef]

58. Iannizzotto, G.; Lo Bello, L. A multilevel modeling approach for online learning and classification of complex trajectories for video surveillance. *Intern. J. Pattern Recognit. Artif. Intell.* **2014**, *28*, 1455009. [[CrossRef](#)]
59. Iannizzotto, G.; Lo Bello, L.; Patti, G. Personal Protection Equipment detection system for embedded devices based on DNN and Fuzzy Logic. *Expert Syst. Appl.* **2021**, *184*, 115447. doi: doi: 10.1016/j.eswa.2021.115447. [[CrossRef](#)]
60. Shuangfeng, L. TensorFlow Lite: On-Device Machine Learning Framework. *J. Comput. Res. Dev.* **2020**, *57*, 1839. doi: 10.7544/issn1000-1239.2020.20200291. [[CrossRef](#)]
61. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*; Curran Associates, Inc.: Sydney, Australia, 2019; pp. 8024–8035.
62. OpenVINO™ Toolkit. Available online: <https://github.com/openvinotoolkit/openvino> (accessed on 11 September 2022).