Published in: [Future Generation Computer Systems] journal.

This version of the article has been accepted for publication, after peer review and is subject to Elsevier terms of use.

The Version of Record is available online at: http://dx.doi.org/10.1016/j.future.2021.07.026

Attention monitoring for synchronous distance learning

Andrea F. Abate, Lucia Cascone, Michele Nappi, Fabio Narducci, Ignazio Passero *

Department of Computer Science, University of Salerno, Via Giovanni Paolo II, 132, Fisciano (SA), Italy

ABSTRACT

In traditional classroom education the teacher can easily perceive or obtain the engagements of the students by observing them. Distance education is affected by the absence of such a feedback coming from expressions and behaviours of the students attending the lesson. Aiming at reducing the gap between these two learning modalities, the proposed system analyses student videos recorded by the cameras available on the laptops by which they are attending the lecture. This approach provides aggregated information on didactic efficacy to the teacher, avoiding the need of sharing student video recordings during the lecture. The teacher is therefore supported during the oral exposition of the lecture. The approach proposed in this study has been conceived as a software architecture running on background and locally on students' personal computers. No sensitive data are shared over the network. It has been evaluated in two experimental sessions dedicated to a sensitivity preliminary evaluation of the proposed instrument and to the assessment of its didactic usefulness by volunteers, students and industrial employees. User evaluation reports, both on student and teacher side, a positive feedback. The discussion can bring to insights and new considerations about learning in general, which is nowadays so significantly forced to change due to COVID-19 pandemic and it could be intended to change even more in the direction of distance learning in the future.

1. Introduction

During a lesson, the teacher requires continuous feedback from students to know if the audience is getting the topic and the explanations provided. In other words, if the lesson is understood and if it is held at an adequate pace. If this is a desirable requirement in classroom education, in case of distance teaching modalities it becomes crucial. The answers to these questions are usually deduced by students by examining their expressions and attitudes, listening to their questions or simply asking them questions directly [1]. However, this approach cannot always be applied both in presence and distance learning (often referred to as online learning also). In a classroom lecture, students may be particularly reluctant to communicate. In distance learning, the absence of a visual feedback from attendees makes the teacher in the condition of not realising the potential efficacy of the lesson for the audience. Currently, the COVID-19 pandemic outbreak forced every level of educational institutes to adopt distance learning as their first choice. Distance learning began in 1850, at the University of Chicago, where for the first time a correspondence program was established in the United States in which the teacher and the student were not in the same physical place. It is a learning method that allows interaction between the teacher and the students remotely without being in a traditional context such as the classroom [2]. In the twentieth century, with the development of new technologies such as radio and television, new ways to spread culture even outside traditional contexts were found. With the advent of the Internet, teaching and learning modalities have changed. New technologies allow for an innovative and fruitful interaction, aimed at overcoming the limitations of past teaching [3]. The full and wide adoption of distance learning has underlined the need for tools to reduce the distance between teacher and students, typical of a remote lessons [4]. Online learning tools, considered as amplifiers of students' participation and reinforcement of their learning have been revealed not stimulating for interaction and attention [5]. Since they turned into the main and solely teaching modality, the adoption of additional/alternative solutions has been perceived as essential worldwide. The aim of this research is to apply Machine Learning (ML) algorithms, already available in the literature, that can offer a measurement of the audience attention by tracing the blinks, gaze and student expressions. Blinking is a good indicator of attention status. When attention level is high, people tend to blink less so as not to lose eye contact with the object of interest. An increase in the blink rate, in contrast, is associated with fatigue. This can best be understood as a cessation of attention inhibition of blinking. The gaze can detect how people acquire information, playing an important role in human communication, reflecting cognitive processes, and also allowing to recognise areas of interest [6]. If the gaze concentrates on a particular area over a prolonged time, the interpretation is twofold: a cognitive difficulty in understanding information or greater interest. Yawning is a mostly involuntary reflex of opening the mouth and inhaling deeply, stretching the muscles of the jaw and trunk. Studies have shown that yawn is not just linked to tiredness, but also to awakening and when there is a change in the state of alertness. The proposed software module analyses these characteristics to be capable of supporting the teacher during a distance lecture. It aggregates information on how the attendee's visual attention is being distributed on the projected slides in the form of heat-maps and provides other aggregated information on the classroom. The teacher can count on a useful tool to overcome the lack of visual feedback in the classroom from the audience and can also understand if a specific topic is confusing the entire audience or a significantly high part of the class by observing the visual patterns that tend to be spread out. The user evaluation study, performed on volunteers from academia and industry, revealed the potentials benefit of such an approach on both actors of a lecture, that are the teachers from one side and the learners from the other. The video processing occurs locally on students' computers thus leading to privacy and security advantages since it prevents the diffusion of sensitive personal situations and also optimises internet bandwidth occupation. The organisation of the paper is as follows. Section 2 provides an overview of the works in the literature that analyse the characteristics mentioned above to infer information on the subject's cognitive state. Section 3 contains the description of the proposed approach and the adopted Machine Learning inferences. We evaluated the quality of our proposed system in terms of sensitivity and didactic benefits in Section 4. Section 5 concludes the discussion by presenting the rationale for such a kind of study and the potentials to be further explored in future research activities.

2. Related works

Today, nearly two billion students around the world have seen their education blocked by the spread of an unexpected pandemic and have been forced to approach new distance learning modalities. In addition to the obvious problems associated with this condition, such as the difficulty of having reliable Internet access and/or the first impact with using new software [7], in [2] the authors reported the main issues perceived by the students with this type of learning, including distractions such as text messages or the use of social networks; boredom and tiredness, boring and unattractive contents that hardly stimulate their attention. In [8] the authors surveyed 578 students in physics. They proposed a statistical analysis to investigate the attitudes, criticalities, and benefits of online learning. The research revealed that good communication abilities and self-organisation skills are positively correlated with perceived learning achievement. On the other hand, students tend to be significantly distracted by the home environment when attending remote lessons. The conclusions of the work suggest offering special courses for promoting self-regulated learning skills, emphasise the positive aspects of distance learning, and (iv) install networking services for supporting student communication. In distance learning the most serious lack from teachers' perspective is the feeling of being heard and understood. In presence, teachers can intercept the student's gaze for immediate feedback. In this point of view, eye tracking systems represent election tools to be used to estimate the student's attention and the cognitive load of solving problems [9,10]. However, asking students to wear an eye tracker at home during a remote lecture is rather unfeasible for a wide collection of reasons. One among all others, a reliable eye tracker is an expensive device. Similarly, Yang et al. [11] develop a brainwave signal-based system to rate students' attention levels as high or low while engaging in learning activity by watching a video lesson. In an e-learning environment, this system could provide timely feedback to online instructors when low attention is registered. Human expressions are the main source of information, along with words, in determining individual's emotions [12]. The heart rate variability (HRV) is another non-invasive marker associated with emotional and cognitive regulation [13]. Several studies show how this measure can be influenced by sleepiness, stress or excessive fatigue [14]. Again, the ones just mentioned are other examples of approaches that take benefits from physical devices. To overcome these limitations and providing a support to teachers during distance learning, different characteristics can be extracted from the analysis of the video stream of a camera acquiring the student frontally while attending the lecture. Several deep learning approaches for evaluating user expressions are available in literature. An example is the approach proposed by Abate et al. [15] that adopts a convolutional neural network for classifying user expressions to dynamically adapt the user interface on user's emotional state. There are various feelings that can be evident and observable from a face, such as sadness, boredom, happiness, confusion and fatigue [16]. A good indicator of fatigue can be yawning. Yawning detection is particularly used to estimate the vigilance level and monitor driver fatigue level showing interesting results [17]. This information associated with excessive blinking or prolonged closing of the eyes can allow obtaining an accurate detection of the state in which the subject from whom this data is extracted is located in terms of fatigue, tiredness, boring or sleepiness. From existing works in the literature it is clear that there is a correlation between a very low blink rate, related to those performances that require high visual attention, and a higher blink rate just before sleep and during boring tasks [18]. Some scholars believe that blinks can provide useful information on central nervous system activation and fatigue levels, as fatigue in completing a task. On student's performance decreases, the blink increases in frequency and the duration. So the changes in these measures appear to be related to changes in vigilance performance [19] and correlated to task demands [20]. Other eye features, such as fixations and saccades, in addition to being useful in detecting emotion [21], can also be a good indicator of attention level. The visual attention of the user can be investigated by analysing those areas on which the gaze lingers for longer or more frequently, the so-called areas of interest (AOI) [22,23]. It can also be studied as a heat-map depicting not single fixation points but sets of points [24]. In [25] the authors analysed the scan-path of several students while they were intent on solving different problems with access to worked examples. Based on the characteristics of shape, direction, length, position and duration, the similarities between the various scanning paths were studied to identify the different attentive pattern during the activity. In [26] the authors present a non-invasive real-time smart system to monitor the attention of each person who is part of a team. The goal is to classify their level of attention, having a preliminary knowledge of the task, and analysing the interaction of each of them with their computer. Also keyboard and mouse biometric behaviours can be analysed to assess students classroom performance while using computers in [27,28]. The authors mainly focused on four different types of classes, such as audio, text, video and image. To evaluate the level of attention of students Machine Learning techniques were used. Distance learning does not deal exclusively with teaching. The COVID-19 pandemic forced several educational institutes to also carry exams and verification in remote mode [29]. In this context, the contribution of biometrics is crucial. Multi-biometric system for continuous student authentication and accessing on cloud platforms have been

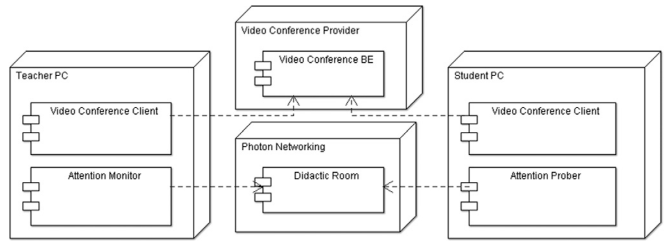


Fig. 1. The deployment diagram of the proposed architecture.

proposed [30–32]. They suggest how face recognition, combined with other physical and behavioural biometric traits, can be analysed in real-time to empower the distance learning modalities, making them more secure and reliable.

3. The proposed approach

The proposed didactic approach combines a regular synchronous online learning instrument, adopted for a video lecture, with a software module running as a service and capable of performing deep inferences on students' videos captured by the laptop camera. In the cases of desktop PCs, the previous setting can be simulated by fixing a common webcam on the top of the PC Monitor. The student application does not expose a graphical user interface (GUI), since just the unidirectional communication from the students to the teacher has been considered at the moment. The teacher is provided with aggregated indicators on the flow of typical oral lecture. Lessons are not all the same. In some cases, the students are significantly asked to take notes during the lesson. In this case, the gaze is forced to switch from the sheets on the desk and the monitor of the personal computer in a rapid and continuous manner. A special condition that is not particularly considered in this study. Rather, the proposed software module considers all those lectures where the attendees are predominantly asked to listen to the lecturer and watch the shared slides. In future development, the system can be improved by establishing an additional channel between the teacher and the students to soliciting attention in the cases of detected distraction.

Fig. 1 shows the architecture of the proposed didactic platform. In the proposed setting, as it is possible to notice, both the teacher and the students adopt a regular Video Conference system for synchronous distance lectures (e.g., Microsoft Teams, Zoom or others). The teacher shares the PowerPoint presentation of the lecture and the students just attend the video lecture. Before starting the lecture, the students launch the "Attention Prober" application and the teacher his/her "Attention Monitor". The "Attention Monitor" aggregates information for the teacher and exposes the heat-map made at run-time with students' gaze directions and the distributions of classified expressions. In particular, the user expressions, shown in the centre-low side of teacher GUI (Fig. 2) are organised in an histogram labelled with expression emoticons. This information quickly signals to the teacher useful hints on the flow of distance didactic action. The applications are connected via Photon[1] networking engine, that, designed as a chatting and multiplayer facility for gaming, supports distance lectures by ideally shifting the "room" metaphor to the "classroom" one (Didactic Room). The teacher arranges the applications on its main screen or puts the game "Attention Monitor" on its second monitor, if available.

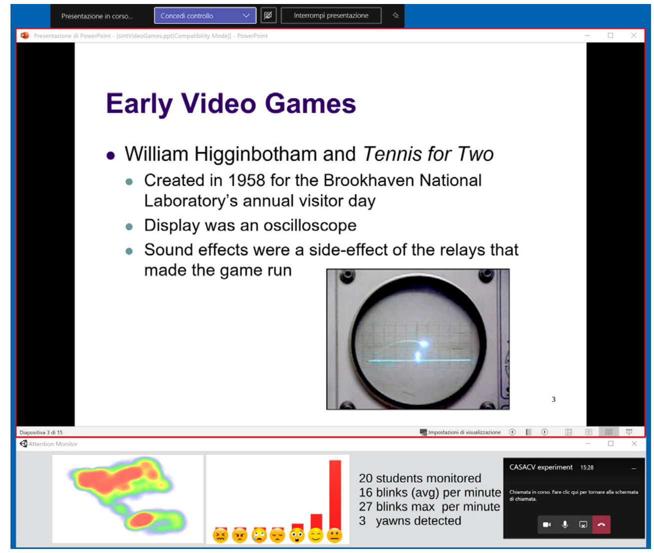


Fig. 2. The Attention Monitor GUI, set by side of teacher Video-lecture software.

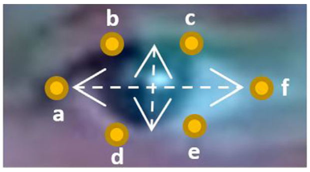
During the controlled experiment Microsoft Teams was adopted as a distance lecture environment, also according to our teaching habits. The lecture has been driven by PowerPoint, with the presentation software set in reading modality. We did not use typical presentation or speaker modalities because these require a full screen presentation that covers the "Attention Monitor" application. As it is possible to notice in Fig. 2, the Teacher GUI consists of a predominant upper portion of the screen that is reserved to the content of the lecture while the bottom is intended to provide the instant feedback about the audience attention in the form of the cumulative heat-map, histogram of user expressions and measurements of blink/yawn rates. From the student's side no any GUI is required since the Attention Prober runs in background accessing the webcam and analysing the video-frames. Moreover, a GUI on student's side could be just a source of distractions from attending the lecture.

3.1. The machine learning approaches

As introduced above, the proposed module exploits computer vision solutions to infer attention information on students attending the lecture. From a real-time acquisition of the camera mounted on the screen, the student's behaviour is analysed through the detection of blinks, yawns and fixations, to provide overall feedback on efficacy of student's attendance to the lecturer. Blinking is a self-regulating, natural and instinctive response, which allows keeping eyes healthy. It is performed periodically with the aim of moistening and cleaning the surface of the eye. The rapid eye movement performed to move a peripheral region to the centre of the visual field is called saccade. Fixation is the time between two saccades and this feature is particularly interesting because it reveals how people acquire information. It is the result of a sophisticated mechanism aimed at stabilising the gaze on the target, in order to allow the analysis of the image from part of the visual system. Four tasks, based on Machine Learning and computer vision, have been implemented in the proposed attention module and consist in:

- *Gaze Tracking*: is in charge of locating the eyes and generating a heat-map of the sight.
- *Blink detection*: a counter of eye blinks during the lecture and the relative frequency.

[1] https://www.photonengine.com/pun.



$$EAR = \frac{||b-d|| + ||c-e||}{2||a-f||}$$

Fig. 3. The calculation of the eye aspect ration (EAR) by which counting the eye blinks.

- *Expression detection*: a ML detector of user expressions during the lecture.
- *Yawn detection*: a counter of yawns during the lecture as a possible indicator of limited attention to the lecture.

The gaze tracker matches the point projected on the screen with the eyes rotation frame-by-frame. It uses face landmarks detection, with a specific focus on eyes, and a calibration process. The results of this module are the generating of a heat-map of the student's attention on the slide broadcast during the lecture. The software module has been implemented on the GazeFlow application[2] and generates heat-maps like those presented in Fig. 5. The calibration procedure consists in a prior step to make the module running properly and reliably. The student is asked to sit in front of the monitor and to follow a moving marker on the slides by the eyes. This process indirectly estimates the distance between the observer and the monitor so that a reliable estimation of the fixation point can be inferred from all possible rotations of the eyes. The module is also slightly invariant to the head pose and to limited translations of the head, which must be necessarily well framed in the camera stream. The estimation of the point observed onto the screen is learned through a simple Convolutional Neural Network (CNN) consisting of a few convolutional layers and a softmax layer. During the calibration process it maps the aspect of the eyes rotations with projected point. The lightweight version of the neural model makes possible to achieve a limited computing demand of the module, a desirable constraint for a module running in background which must not impact on student's lecture attendance.

Automatic detectors of eye blinks have been proposed for many tasks including attention level estimation [33]. In general, the frequency of the eye blinks can be strictly related to cognitive activity. For this reason, the proposed attention module implements a blinking counter[3] which is based on the work of Suokupova and Cech [34]. The detection of blinking is a simple image processing tasks. Starting from the eye landmarks, the eye aspect ratio (EAR, Fig. 3) to count the number of blinks can be computed. It consists in the following equation:

The useful properties that can be extracted from eye blinks are: (i) the blink rate, (ii) the duration, and (iii) the amplitude. The blink rate is the number of blink per minute which value is about 17 blinks/minute on average [35]. The blink duration, meaning the time interval between the blink starting and ending ranging from 60 ms to 700 ms, is another good indicator of the mental state of the subject [36]. Blink amplitude, which represents the level of openness of the eyes, is a very variable measurement that may differs significantly among people. Considering the operating condition in this study, the blink detector processes videos acquired by personal devices. Such devices can differ significantly to each other both in quality and frames per second. For this reason, just the blinking rate has been considered.

Similar to blink detectors, the landmarks detected in the labial area of the face can be processed to detect the expressions[4] and yawns.[5] Yawns are a special class of expression classification problem commonly associated with drowsiness. In many cases, the rate of yawns is a good indicator of lack of attention in the monitored person (it is usually taken into consideration for estimating the level of attention of drivers while driving a car [37]). Computing the distances between landmarks on the upper and lower lips, the yawns can be easily detected by thresholding the measurements in real-time at the same way the blinks are detected. The yawn detector is based on the assumption that a yawn causes a strong and prolonged opening of the lips. Computing yawn duration and amplitude at the same way as for eye blinks, the dynamics of the yawn can be effectively distinguished by others like speaking or laughing (refer to Fig. 4). Therefore, a counter of the yawns can be simply obtained once labial landmarks have been detected.

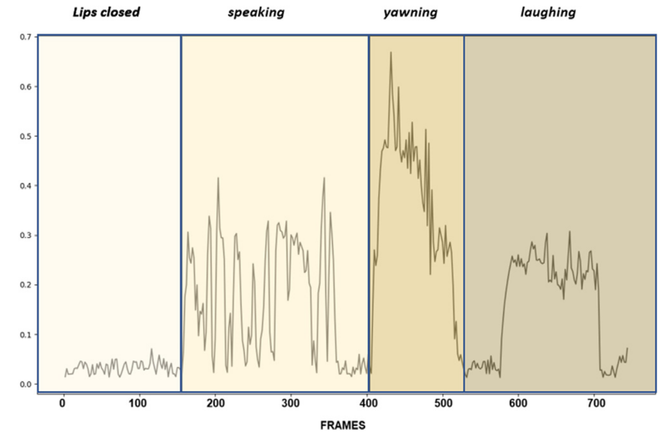


Fig. 4. Plot showing the difference in amplitude and duration of yawns with respect to other common tasks that can be inferred from labial landmark dynamics.

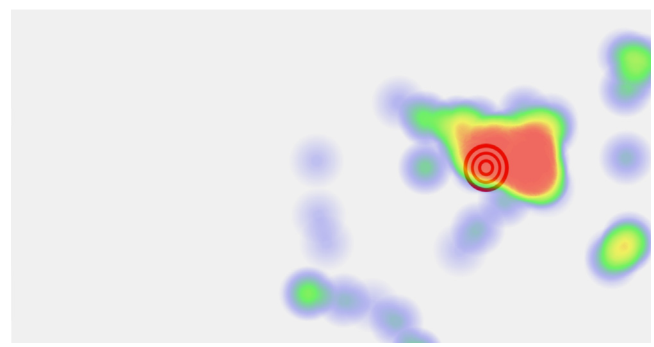


Fig. 5. The image shows one of the heat-map obtained during the calibration process superimposed on the slide projected.

3.2. Calibration of the algorithms

Aiming at tuning the adopted ML algorithms, a preliminary training session has been performed. The tuning was performed for customising the detection parameters with respect to camera and to the different point of views. Moreover, it were also meant to provide a first evaluation of the sensitivity of the proposed system on gaze detection applied to the students' video recordings.

The calibration has been performed on the videos recorded by the students during a synthetic distance lecture. The teacher

2 https://github.com/szydej/GazeFlowAPI.

3 https://www.pyimagesearch.com/2017/04/24/eye-blink-detection-opencv-python-dlib/.

4 https://github.com/pablovin/FaceChannel.

5 https://github.com/kostasthanos/Drowsiness-Detection.

Table 1
The pre-experiment questionnaire.

Question	Score				
q1. I like playing video games	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
q2. I spend one hour a day playing video games	Always	Often	Sometimes	Rarely	Never
q3. I buy one or more video games a month	Always	Often	Sometimes	Rarely	Never
q4. I like retro gaming	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
q5. I have one or more video games	>3	3	2	1	0
q6. I will buy a video game in the next month	Definitely	Probably	Possibly	Probably not	Definitely not
q7. I like reading books (not school ones)	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
q8. I read a book	Always	Often	Sometimes	Rarely	Never
q9. I buy one or more books a month	Always	Often	Sometimes	Rarely	Never
q10. I like old books	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
q11. I have one or more books	>3	3	2	1	0
q12. I will buy a book in the next month	Definitely	Probably	Possibly	Probably not	Definitely not

asked to follow an 'ad hoc' presentation designed to solicit the students to observe known points of the slide, represented by specific targets. In this way, it has been possible to adapt the detector to differences in camera lenses in numerous settings and to evaluate if the students' videos are suitable for the analysis to be performed. Additionally, the calibration required the students to mimic some expressions aiming at evaluating the sensitivity of Machine Learning algorithms involved in the proposed processing pipeline (more details can be found in Abate et al. [15]).

We asked the participants to record their videos sitting on the chair at a distance of about 60–70 centimetres from the monitor, and consequently from the camera. In the second session, we required students to stay closer at 30 centimetres from the camera (see Fig. 6). This to be able to have an upper bound of the quality of the videos that can be currently recorded via laptop cameras, even if this distance is not practical to maintain during a real remote lecture and serving just as a benchmark. The tuning has been performed by adopting an aimed synthetic PowerPoint presentation designed to solicit student gazes in specific and known directions: the students have been explicitly asked to follow a marker positioned at several known positions on the slides. In this way, it was possible to take into account all the differences in physical settings adopted by the students. At the end of the calibration, we found that 60–70 centimetres (i.e., the regular distance from laptop screen) is a good 'point of observation' on students and it well balances the image resolution and gaze detection.

Fig. 6. Samples of participants during the calibration process. From left to right, the first column shows the participants while pointing at the moving marker on the screen by just rotating their eyes at a close distance from the screen. The second column is the same condition as the first one but at a higher distance from the screen. The third and fourth shows participant while rotating their heads to follow the marker at a close and far distance from the screen.

4. Evaluation

The support provided by the proposed approach to synchronous distance learning has been evaluated in two controlled experiments: the one performed with the voluntary participation of 20 students attending the "Context Aware Security Analytics in Computer Vision" (Computer Science Master Degree, Internet of Things Curriculum of the University of Salerno, Italy) and the other involving 30 computer scientists and engineers of Kineton, an engineering company specialised in the automotive, media, entertainment, and telecommunications sectors. It is important to point out that the topics adopted for this assessment are not part of the student course program (as well as are not related to working practices for Kineton employees) and that, neither the students nor the employees, have been evaluated during the experiment. In addition, as it is possible to deduce by the system description, the tool is specifically formulated for Distance Education actions and the experiment has been performed at distance in safe conditions, without exposing the participants to pandemic risks. After the tuning of the applications, the proposed system has been evaluated in two sessions, that are Didactic session with students and Industrial session with employees. Both the sessions have been focused on the application of the proposed analysis to the video captured during two oral presentations performed by the teacher in distance learning modality. In particular, the same two lectures have been adopted for both the sessions and were focused on the "History of Video Games" and on the "History of Typography". The idea behind the choice of the two topics was to propose two lectures that should evidently receive different attention from the participants. Indeed, because of being computer scientists in both the Didactic and the Industrial sessions, the involved volunteers clearly should prefer video games with respect to typography and we were expecting to detect this preference in evaluation results.

The rationale behind this choice is that we expect more interest in topics about the video games by an audience of computer scientists, which should be reflected in the behaviours of the attendees during the lessons in terms of gaze, blinks and yawns.

As suggested in [38], we adopted a fully balanced design for the experiment in both the evaluation sessions: half of the participants started with video games lecture, and the other subjects with typography one. This prevents from fatigue or boredom to bias the results. Before the controlled experiment, a pre-evaluation questionnaire has been submitted to participants to control their degree of interest toward the presented topics. The questionnaire reported in Table 1 assesses the attitudes and practices of the participants by direct 12 questions (6 for topic) but also in terms of purchase intentions. The Pre-Experiment questionnaire establishes a classification of user attitudes toward Video Games and Typography in general. The twelve questions are organised as follows: questions from q1 to q6, are focused on Video Games while their counterparts for Typography are the questions from q7 to q12. According to their nature, the questions require different categorical answers.

Table 2
The post-experiment questionnaire.

The video games and typography questionnaires	
Q1	I was interested in the presentation
Q2	I found the presentation well organised
Q3	I was concentrated on the teacher speaking
Q4	I found teacher exposition to be clear
Q5	I was concentrated on the slide flow
Q6	I had troubles attending lecture

Table 3
Hypothesis testing on the pre-experiment results.

Pre-experiment wilcoxon signed rank test				
Test	h_0	p-value	Med_V	Med_T
q1 vs. q7	1	0.0005	2	3
q2 vs. q8	1	0.0172	2	2
q3 vs. q9	0	0.08	2	3
q4 vs. q10	1	0.015	2	4
q5 vs. q11	1	0.0006	1	3
q6 vs. q12	1	0.037	3	3

After every presentation, the participants were asked to fill the "Post-Experiment Questionnaire" reported in Table 2, with answers anchored on the 5 points Likert scale going from "Strongly Agree" (anchored to 1) to "Strongly Disagree" (anchored to 5). The questionnaire aims at subjectively evaluating the expected interest toward the two subjects and the concentration applied during the attended distance lectures.

Both the controlled experiments have been articulated in the same two phases:

- the preliminary ethnographic survey aiming at assessing the characteristics of the participants to the experiment (Pre-Experiment Questionnaire);
- the on-the-field evaluation performed between the two different degrees of participant's interest perceived and measured during the two distance lectures (video games and typography).

Both the presented lectures adopted in the last phases are made of 14 content slides and one header slide: the oral presentations performed by the teacher lasted both around 14 min, one minute per slide. The contents are simple and minimalist: just black text on white background with the addition of equal number of images (a sample is shows in Fig. 2).

The controlled experiment checks the participant's interest in the two topics proposed by asking him/her to answer a questionnaire before starting the two didactic sessions (first phase), and by directly asking the opinions after both the presentations. Objective measurement of user attention was performed, off-line, by measuring gaze fixations expressed as a percentage of the slide duration (see Section 3.1). The controlled experiment has provided two classes of results, with respect to didactic and industrial users:

- a subjective evaluation of participant's interest (before and after the presentations) toward the two lectures, collected via questionnaires.
- objective measurements collected by analysing students' gaze direction (also in terms of fixations), their expressions, and counting detected blinks and yawns.

At the aim of this evaluation, only gaze direction has been taken into consideration, utilising fixation times as a rough estimation of student attention. Other metrics that can complement user behaviour analysis, considering also slide area coverage, may be introduced in future developments of the presented research.

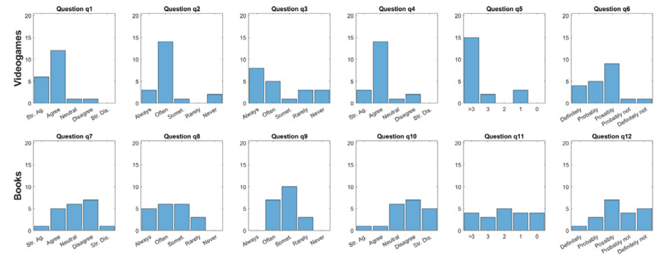


Fig. 7. The results of Pre-Experiment questionnaire.

4.1. The didactic session

The Pre-Experiment Questionnaire, reported in Table 1, has been answered by all the participants (Didactic and Industrial) before starting the two lecture sessions.

As it is possible to notice from results in Fig. 7, Computer Science students participating in the experimental sessions (age ranging in 21–25) confirmed the assumption that they are more interested in video games than in a lecture on the history of typography. This is evident by analysing the answers to the Pre-Experiment questionnaire. By comparing the first six answers (Video Games) with the last ones (Typography), it is possible to have a preliminary estimation of the expected levels of attention the students will pay to the two lectures.

The differences in users' attention and interest toward the adopted topics have formally been evaluated in pairwise statistical hypotheses verification tests. In particular, since the sample dimension did not let us choose a parametric test, we adopted the Wilcoxon Signed Rank test [39], that evaluates the null hypothesis h_0 of equal medians and does not impose normality assumptions. Table 3 reports the results obtained by the pairwise comparison performed on the associated questions. The table reports the tested couple in the column labelled "Test", in the column labelled with 'h_0', the result of the test (0 or 1) and the p-value. The two rightmost columns report the answer medians: Med_V for Video Games and Med_T for Typography. As it is possible to notice, the majority of the comparisons does not statistically support the null hypothesis that the two samples are characterised by equal medians at the 0.05 level of significance. Two comparisons do not highlight statistical differences: the q2–q8 and q3–q9 tests. In these cases, the formulation of the questions, the first concentrated on effective gaming/reading time spent a day and the second one on video games/books purchase, are probably too practical to be a good indicator of the expected different user preferences.

4.1.1. The didactic session results

After the Pre-Experiment questionnaire, we were expecting to detect the same differences in the answers given to the Post-Experiment questionnaire, shown in Table 2. In particular, we aimed at showing the same differences underlined by the previous phase of the evaluation, on objective indicators obtained by Deep Learning techniques applied to the student videos.

The results of the Post-Experiment questionnaire are aggregated in the histograms depicted in Fig. 8. The upper line is for Video Games while the lower one is for Typography. The formal evaluation of participants' answers collected via the Post-Experiment questionnaire, has been performed with 6 Wilcoxon Signed Rank test at the 0.05 level of significance [40]. Concerning the questions, reported in Table 2, it is important to point out that Q1, Q3, and Q5, analogous for both the topics, are the 3 questions assessing the effective concentration of students during the didactic experience. Questions Q2, Q4, and Q6 have just a

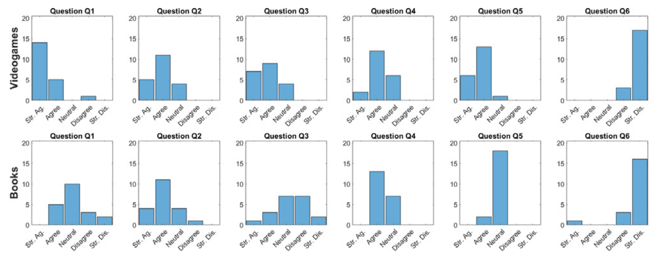


Fig. 8. The results of the two Post-Experiment questionnaires after Classroom evaluation.

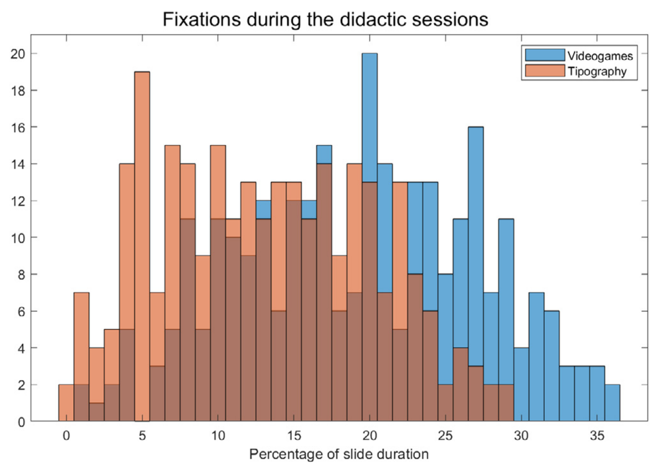


Fig. 9. The fixations of students' gazes expressed as the percentage of slides' duration.

control role, instead, and they aim at excluding bias factors like PowerPoint slide organisation, teacher exposition style or other kinds of problems that may arise during a distance lecture. In the case of the Post-Experiment questionnaire, the comparison of Q1, Q3, and Q5 questions underlines a significant statistical difference in participants' interest toward the two topics proposed. As we would have expected it to happen, the control questions Q2, Q4, and Q6 do not support, with sufficient statistical evidence, the presence of problems in the organisation and fruition of the two lectures (see Table 4).

As shown in the previous sections, the ML algorithms applied to student images provide several hints to the teacher about lecture flow in terms of student's gaze direction with respect to the slide, expressions, and yawns. At the aims of the evaluation, we adopted the fixation metric. The idea is to record students' gaze fixations expressed as a percentage of the entire time the slide has been presented.

Fig. 9 depicts the histograms of collected fixation times aggregated by session.

The average fixation time collected during the Video Games lecture is 19.31% and the one related to Typography is 13.1%. As it is possible to graphically notice, is evident the difference in fixation times between the two topics presented.

Before performing a formal hypothesis verification test to support the previous observation, we executed a Lilliefors test [41] on the measured fixation times. The test checks the null hypothesis that the data are normally distributed. Applied to both the fixation times measured for Video Games lesson and for Typography one, the Lilliefors tests did not support the null hypothesis of normality. Accordingly, we evaluated the differences between the two groups of results by applying Wilcoxon Rank Sum test at the 5% of significance and paired on the slides.

Table 4
Hypothesis testing on the classroom post-experiment results.

Post-experiment wilcoxon signed rank test				
Videog. vs. Typogr.	h_0	p-value	Med_V	Med_T
Q1	1	0.0004	1	3
Q2	0	0.59	2	2
Q3	1	0.0006	2	3
Q4	0	0.37	2	2
Q5	1	0.0002	2	3
Q6	0	0.76	5	5

Table 5
Classroom session: Gaze fixation times comparison aggregated by slide number.

		Rank sum	p-value	Med_V - Med_T
Slide	1	1	0.0007	14
	2	0	0.1009	3.5
	3	0	0.0674	6
	4	1	0.0131	3.5
	5	0	0.1224	2.5
	6	0	0.1888	0.5
	7	1	0.0254	4
	8	1	0.0070	10
	9	1	0.0052	6.5
	10	1	0.0003	10
	11	1	0.0113	7
	12	1	0.0142	5
	13	0	0.5880	2.5
	14	1	0.0036	9

Table 5 reports the results of the 14 Wilcoxon Rank Sum tests performed comparing the detected fixation (times reported in percentage of slide duration) on the corresponding slides of the two lectures proposed. The rightmost column of the table shows the differences in median fixation values (relative to all participants) for every slide.

Table 5 shows that only in the case of five slides (2, 3, 5, 6, and 13) the tests did not find enough statistical evidence to reject the null hypothesis of median equality. It is important to point out that, also, in this case, the punctual statistics reported in the rightmost column confirm the trend of the other rejecting 9 tests. After the Didactic Session, we can conclude that the objective evaluation performed on the ML "raw" indicator of participants' attention (Fixation time) has confirmed the results obtained by the questionnaires answered by the participants after the experiment. We believe this assonance as a good indicator of the efficacy of the proposed approach. In addition, also in the cases of missing statistical evidence, the positive differences of medians (see Table 5) do not provide elements against the previous findings.

4.2. The industrial session

The second evaluation session performed has involved 30 voluntaries working at several technical degrees in a engineering and software development company. Their education ranges among bachelor and master degrees in computer science and electronic engineering. Also in terms of age, this sample is more heterogeneous with respect to the one participating to the Didactic Session, even if the ages are in the range 24–37. The Industrial session adopts the same design of the Didactic one, and has begun with the Pre-Experiment Questionnaire (Table 1). As it is possible to notice from the results depicted in Fig. 10, the expectation about the propensity towards the proposed topics are confirmed also for the employees participating in the industrial experimental sessions: because of their technical skill and of their attitudes, also the second group of participants appears more interested in video games than to the history of typography. This is

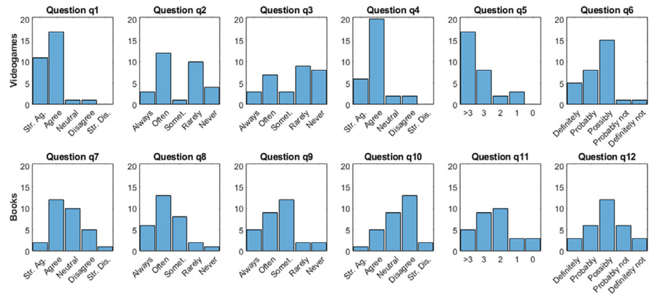


Fig. 10. The results of the Pre-Experiment questionnaire for the Industrial evaluation.

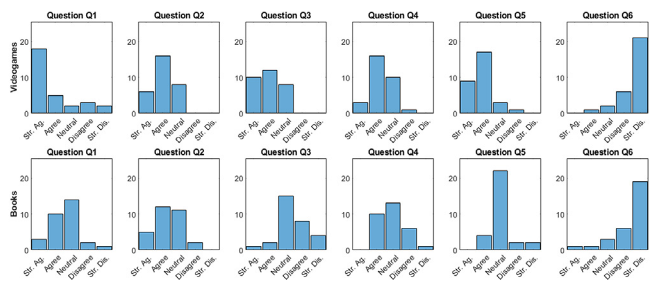


Fig. 11. The results of the two Post-Experiment questionnaires after the Industrial evaluation.

even more evident, compared to the Didactic session participants, by analysing the distribution of answers to the Pre-Experiment questionnaire.

Also in the case of the Industrial Session, the differences in interest toward the adopted topics and in user attention have been formally evaluated with 6 pairwise Wilcoxon Signed Rank tests [39]. Table 6 (the structure is detailed in Section 4.1) reports the results obtained by the pairwise comparison performed on the associated questions.

In accordance with the graphical observations reported before, in the Industrial session, only the q6–q12 comparison is not supported by statistical differences. Aiming at replicating the same conditions of the Didactic session, the participants to the Industrial session have attended the recordings of the two presentations focused on the history of Video Games and of Typography. Also in this case we adopted a fully balanced order: half of the participants began the experiment with the Video Games presentation and the others with the Typography one.

4.2.1. *The industrial session results*

Also in the case of this Session, we detected the same differences in the answers given to the Post-Experiment questionnaire, submitted to participants immediately after each distance lecture. The results are shown in Table 7. Collecting the ML analyses of the participant, we aimed at detecting the same differences on objective indicators.

Fig. 11 organises the results of the Post-Experiment questionnaire in histograms. The upper line is for Video Games while the lower one is for Typography. Also in the case of Industrial Session, the formal evaluation of participants' answers to the Post-Experiment questionnaire, has been performed with 6 Wilcoxon Signed Rank test at the 0.05 level of significance [40]. In the case of the Post-Experiment questionnaire, the comparison of Q1, Q3, and Q5 questions underlines a significant statistical difference in participants' interest toward the two topics proposed. As we would have expected it to happen, the control questions Q2, Q4, and Q6 do not support, with sufficient statistical evidence,

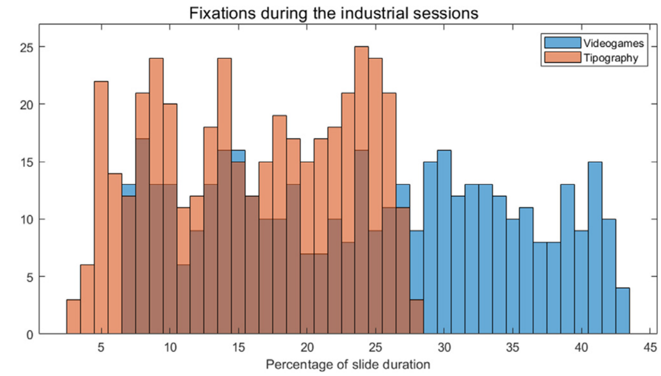


Fig. 12. The fixations of employees' gazes expressed as the percentage of slides' duration.

Table 6
Hypothesis testing on the industrial pre-experiment results.

Pre-experiment wilcoxon signed rank test				
Test	h_0	p-value	Med_V	Med_T
q1 vs. q7	1	5×10^{-5}	2	3
q2 vs. q8	1	0.047	2.5	2
q3 vs. q9	1	0.0029	4	3
q4 vs. q10	1	4.2×10^{-5}	2	3.5
q5 vs. q11	1	0.0052	1	3
q6 vs. q12	0	0.164	3	3

Table 7
Hypothesis testing on the industrial post-experiment results.

Post-experiment wilcoxon signed rank test				
Videog. vs. Typogr.	h_0	p-value	Med_V	Med_T
Q1	1	0.0169	1	3
Q2	0	0.1719	2	2
Q3	1	1.8×10^{-5}	2	3
Q4	0	0.0076	2	3
Q5	1	1.4×10^{-5}	2	3
Q6	0	0.502	5	5

the presence of problems in the organisation and fruition of the two lectures. Also the medians of scores, reported in the table, support this founding.

Respecting the design of the Didactic Session, also for the industrial evaluation, we adopted the fixation expressed as a percentage of the entire time the slide has been presented.

Fig. 12 reports the histograms of the fixation times aggregated by session. In the case of the Industrial Session, the average fixation times collected are 24.31% for Video Games lecture and 16.1% for the one related to Typography. As it is possible to graphically notice, the difference in detected fixation times between the two topics is evident in the figure and it is higher than the one reported in the Didactic Session. We believe that the students are more accustomed to being concentrated on a presentation they are not interested in, with respect to the employees.

After the Lilliefors test [41], we evaluated the differences between the two groups of results by applying 14 Wilcoxon Rank Sum tests at a 5% of significance. Table 8 reports the results of the tests. Also in this case, in the case of five slides (3, 4, 8, 10 and 11) the tests did not find enough statistical evidence to reject the null hypothesis of median equality. However, also for the Industrial Session, the punctual statistics reported in the rightmost column confirm the trend of the other 9 tests. The objective evaluation performed on the ML "raw" indicator of participants' attention (Fixation time) has confirmed the results obtained by the questionnaires answered by the participants after

Table 8
The industrial session: Gaze fixation times comparison, aggregated by slide number.

		Rank sum	p-value	Med_V - Med_T
Slide	1	1	0.019	10
	2	1	0.014	8
	3	0	0.069	8
	4	0	0.23	9
	5	1	0.045	9
	6	1	0.0178	6.5
	7	1	0.0012	4.5
	8	0	0.068	8
	9	1	0.003	6
	10	0	0.29	6.5
	11	0	0.3	4.5
	12	1	0.0043	16.5
	13	1	0.0099	10.5
	14	1	0.0093	5.5

the experiment. We believe this assonance to be a good indicator of the efficacy of the proposed approach.

4.3. Threats to validity

In this subsection, we detail how we designed and performed the experiment to maximise, to the best of our possibilities, the validity and generalisation of the results. We investigate threats that may impact the validity of the results according to the classification proposed by Wohlin [38] and organise the discussion with respect to internal, external and conclusion validity. Concerning the internal validity, investigating eventual influences that can affect the independent variables with respect to causality, it is important to remark that aiming at avoiding learning effects due to presentation sequence, the experiment was organised following the fully balanced design with an equal number of students starting with each topic. External validity describes the study representativeness and the ability to generalise the results outside the context of the evaluation. We identified the following threats to external validity: (i) the subject population we selected was composed of 20 Computer Science students and 30 employees of an engineering company, that are particularly favourable to the adopted technologies, (ii) the experimentation should be extended to students of other subjects, (iii) the age of subjects is only partially representative respect to the potential applications of the proposed system and the evaluation may be enriched with experiments performed on other categories of potential users like adults or elderly. However, we have selected two participant samples that slightly differ in some characteristics thus contributing to the exclusion of possible bias factors. Conclusion validity examines the issues affecting the ability to draw the correct observations. The participant number is overall 50, but the heterogeneity of the sample limits the collection of normally distributed data. In fact, after normality check, we adopted non-parametric Wilcoxon Signed Rank and Rank Sum tests and we respected the requirements providing good statistical power for the validity of the results. However, the performed evaluation is only aimed at inspiring considerations about how the subjective behaviour is related to objective measurements adopted, and to such a purpose we performed formal experimentation.

4.4. Teacher impressions

In the last year, pushed by the urgency of adopting distance didactic, almost all teachers have taken their lectures in a distanced manner. The common teacher opinion about distance teaching is the lack of feedback about the flow of the lecture, about teacher exposition style and on the arrangement of the concepts. The feeling is to be in front of a rubber wall and the teacher exposition rarely gives rise to questions or interaction with the students. The teacher involved in the experiment has greatly appreciated the support provided by the proposed tool. Without being too intrusive, the didactic analyser proposed provides interesting hints to the teacher and lets him/her be able to adapt the flow of the oral exposition to student attention and to improve the organisation of the course by exploiting the continuous hints provided by the student gaze and expression analysis. The instrument has been proved effective also in an evaluation session performed in the context of a private company that is often involved, like other companies, both similar and operating in different commercial sectors, in a remote training session for professional updating of the staff. What is still required by the teacher is a feedback channel enabled toward the students who are not attentive to the lesson. Currently, the teacher directly speaks to the classroom and may solicit the students when he/she sees the attention of the audience to decrease. The implementation of an automatic system for improving student attention, may keep them focused on the lecture and may enable an attention checking/soliciting questions in the cases of possible fall of concentration.

5. Conclusion

The continuous improvement of network bandwidth and the availability of reliable online services have created the right premises to promote the distance learning as an effective alternative solution to traditional classroom instruction. Online education has several benefits for attendees. It can provide a broad method of communication and can expand education to a wider audience, due to the flexibility and a custom management of time by offline attendance. The advantages are numerous as well as numerous are the instruction institutes that adopted online education as exclusive or complementary means of teaching. Nowadays, COVID-19 pandemic has turned the advantages of online education into a concrete and essential need. Risks of transmission of the virus among the populations led the government to adopt strict regulations to avoid crowds. University and school courses have been severely invested by the restriction, forcing students to attend lectures at home for months (even for one entire year in some countries). The side effect of such a massive usage of online platforms for education has been the collapse of the transmission and efficiency of the networks due to the intensive use of bandwidth to allow teachers and students being visually connected. This results in reduced participant attention and difficulty monitoring students for the lectures during a lesson. While providing a lecture, the lectures can correct the tone of the voice or stressing a particular content of the lesson, according to the visual feedback collected by the audience. Moving to fully online education scenarios, this useful insight misses to teachers. For the motivations explained above, it is unfeasible to think that all students may turn on and share their online recordings of their camera. In this work, a monitoring software module, running locally in background on student's personal computer, is proposed. It consists in a support tool for teachers to gather statistics on students' engagement and attention through Machine Learning and computer vision techniques. By the video stream recorded during the lecture, the software module analyses the attention of the students, monitoring his/her expressions, the eye blinks, the yawns and also tracking the gaze to estimate the fixations of the attention and the scan-path over the projected slide. Exploiting less demanding Machine Learning solutions to computer vision tasks, the software module can run in real-time without interrupting the student attending the remote lecture and avoiding the transmission of sensitive data over the network, which differently should be encrypted with a further overhead. The module provides the lecturer with an overall feedback about the proficiency of the lecture performing inference on students data. The visual feedback consists of a combination of a cumulative heat-map and quantitative measurements of blinks and yawns that allow the teacher to do rapid considerations about the current projected lecture content and possibly solving the issues so to catch the attention of the audience again, without asking the teacher to interrupt the lesson. This is perceived as an added value in distance learning by the attendees who often report a less fluidity of the lecture in remote condition respect to traditional classroom one. The experimental results on volunteers, 20 students and 30 employees, reported positive feedback, both in terms of gaze tracking and evaluation questionnaire. From the student's point of view, the proposed software module is considered non-intrusive and they report a sense of trust due to the chance of not sharing personal recordings. From the teacher side, the proposed software has been considered as a valid support tool to control the flow of the lecture. The teacher is solicited to ask questions when indicators suggested a decrease of attention in the audience, thus filling that gap in the communication that is often sensibly driven by body language and natural attitudes or the experience. In this work, the focus was providing the teacher with feedback during the lecture with the objective of establishing new way of connecting the lecturer with the attendees. The experimental results collected during a user-evaluation, involving students and employees in industrial field, demonstrated the potentials and the limitations of the proposed work. The results achieved in this study can be considered as a preliminary feedback for teachers to assess the efficacy of a lecture and the potential learning rate of attendees. The statistics collected by the software module during the lecture have been considered useful from teacher side to suggest contents of the lecture that resulted harder do understand or distracting the attention of the audience. The teacher can propose recap questionnaire at the end of the lecture that can better stimulate the students to focus on topics that they did not get properly. However, aimed studies are necessary to infer the efficacy of the attention. How the fixations, the eye blinks or the eye scan-path impact on the understanding of the lecture involves a complex user evaluation which is out of the scope of this work, even if the proposed approach represents a preliminary essential step to achieve such a goal. Similarly, including other users categories, like adults or even elder people, into the validation of the approach could provide further insights on social-economic variables influencing the efficacy of the tool. We believe, however, that such an experimentation does not fall into the goals of our proposed work. As another further improvement of this study, the same protocol can be feasibly implemented in distance exam monitoring. Attempts of cheating during exams can sometimes be difficult to detect. In the era of COVID-19 pandemic, which has dramatically changed practices and protocols, these difficulties have been even more stressed. By properly collecting data during exam sessions, Machine Learning algorithms can be trained to detect anomalies thus providing a probabilistic estimate of the reliability of the mark assigned to a written text or the reliability of an oral discussion.

CRediT authorship contribution statement

Andrea F. Abate: Investigation, Formal analysis, Supervision, Visualization. **Lucia Cascone:** Methodology, Investigation, Formal analysis, Writing - original draft, Writing - review & editing. **Michele Nappi:** Investigation, Formal analysis, Supervision, Visualization. **Fabio Narducci:** Investigation, Data curation, Formal analysis, Writing - original draft, Writing - review & editing. **Ignazio Passero:** Investigation, Data curation, Software, Formal analysis, Validation, Writing - original draft, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are grateful to Gian Piero Fiengo and to the engineers and developers of the Kineton Company for participating in the evaluation, and to the students who also volunteered as participants to the experimental sessions.

References

[1] T. Zarraonandia, R. Francese, I. Passero, P. Díaz, G. Tortora, Augmented lectures around the corner? Br. J. Educ. Technol. 42 (4) (2011) E76–E78.

[2] K. Bataineh, M. Atoum, L. Alsmadi, M. Shikhali, A silver lining of coronavirus: Jordanian universities turn to distance education, Int. J. Inf. Commun. Technol. Educ. 17 (2) (2021) 1–11.

[3] O. Bent, P. Dey, K. Weldemariam, M. Mohania, Modeling user behavior data in systems of engagement, Future Gener. Comput. Syst. 68 (2017) 456–464.

[4] J. Song, Y. Zhang, K. Duan, M. Shamim Hossain, S. Rahman, TOLA: Topic-oriented learning assistance based on cyber-physical system and big data, Future Gener. Comput. Syst. 75 (2017) 200–205.

[5] M. Cantabella, R. Martínez-España, B. Ayuso, J. Yáñez, A. Muñoz, Analysis of student behavior in learning management systems through a big data framework, Future Gener. Comput. Syst. 90 (2019) 262–272.

[6] A. Casanova, L. Cascone, A. Castiglione, M. Nappi, C. Pero, Eye-movement and touch dynamics: A proposed approach for activity recognition of a web user, in: 2019 15th International Conference on Signal-Image Technology Internet-Based Systems (SITIS), 2019, pp. 719–724, http://dx.doi.org/10.1109/SITIS.2019.00117.

[7] N. Dietrich, K. Kentheswaran, A. Ahmadi, J. Teychené, Y. Bessiere, S. Alfenore, S. Laborie, D. Bastoul, K. Loubière, C. Guigui, M. Spérandio, L. Tiruta-Barna, P. Etienne, C. Cabassud, A. Line, G. Hébrard, Attempts, successes, and failures of distance learning in the time of COVID-19, J. Chem. Educ. 97 (2020) http://dx.doi.org/10.1021/acs.jchemed.0c00717.

[8] P. Klein, L. Ivanjek, M.N. Dahlkemper, K. Jeličić, M.-A. Geyer, S. Küchemann, A. Susac, Studying physics during the COVID-19 pandemic: Student assessments of learning achievement, perceived effectiveness of online recitations, and online laboratories, Phys. Rev. Phys. Educ. Res. 17 (1) (2021) 010117.

[9] R. Rosiek, M. Sajka, Eyetracking in research on physics education, in: Key Competences in Physics Teaching and Learning, Springer, 2017, pp. 67–77.

[10] S.J. Hansen, J.R. VandenPlas, Eye tracking in chemistry education research: Study logistics, in: Eye Tracking for the Chemistry Education Researcher, ACS Publications, 2018, pp. 11–23.

[11] S. Yang, C. Chen, C. Yu, Assessing the attention levels of students by using a novel attention aware system based on brainwave signals, in: 2015 IIAI 4th International Congress on Advanced Applied Informatics, 2015, pp. 379–384.

[12] C. Bisogni, C. Pero, IFEPE: On the impact of facial expression in head pose estimation, in: International Workshop on Biometric Data Analysis and Forensics IWBDAF.

[13] H. Melo, A. Hoeller, R. Walz, E. Takase, Resting cardiac vagal tone is associated with long-term frustration level of mental workload: Ultra-short term recording reliability, Appl. Psychophysiol. Biofeedback 45 (1) (2020).

[14] F. Rundo, S. Conoci, F. Trenta, S. Battiato, Car-driver drowsiness monitoring by multi-layers deep learning framework and motion analysis, Lect. Notes Electr. Eng. 629 (2020) 169–175.

[15] A.F. Abate, A. Castiglione, M. Nappi, I. Passero, DELEX: A deep learning emotive experience: Investigating empathic HCI, in: Proceedings of the International Conference on Advanced Visual Interfaces, in: AVI '20, Association for Computing Machinery, New York, NY, USA, 2020, http://dx.doi.org/10.1145/3399715.3399820.

[16] S. Umer, R.K. Rout, C. Pero, M. Nappi, Facial expression recognition with trade-offs between data augmentation and deep learning features, J. Ambient Intell. Humaniz. Comput. 1–15.

[17] M. Knapik, B. Cyganek, Driver's fatigue recognition based on yawn detection in thermal images, Neurocomputing 338 (2019) 274–292.

[18] R. Martins, J. Carvalho, Eye blinking as an indicator of fatigue and mental load – a systematic review, 2015, http://dx.doi.org/10.1201/b18042-48.

[19] Detection of vigilance performance using eye blinks, Applied Ergon. 45 (2, Part B) (2014) 354–362, http://dx.doi.org/10.1016/j.apergo.2013.04.020.
[20] A. Magliacano, S. Fiorenza, A. Estraneo, L. Trojano, Eye blink rate increases as a function of cognitive load during an auditory oddball paradigm, Neurosci. Lett. 736 (2020).
[21] J. Przybyło, E. Kańtoch, P. Augustyniak, Eyetracking-based assessment of affect-related decay of human performance in visual tasks, Future Gener. Comput. Syst. 92 (2019) 504–515.
[22] T. Shotton, J. Kim, Assessing differences on eye fixations by attention levels in an assembly environment, Adv. Intell. Syst. Comput. 1201 AISC (2021) 417–423.
[23] S. Cullipher, J.R. VandenPlas, Using fixations to measure attention, in: Eye Tracking for the Chemistry Education Researcher, ACS Publications, 2018, pp. 53–72.
[24] M. Borys, M. Plechawska-Wójcik, Eye-tracking metrics in perception and visual attention research, EJMT 3 (2017) 11–23.
[25] S. Stranc, K. Muldner, Scanpath analysis of student attention during problem solving with worked examples, in: International Conference on Artificial Intelligence in Education, Springer, 2020, pp. 306–311.
[26] D. Duraes, D. Carneiro, J. Bajo, P. Novais, Modelling a smart environment for nonintrusive analysis of attention in the workplace, Expert Syst. 35 (2018) http://dx.doi.org/10.1111/exsy.12275.
[27] D. Durães, Student attention evaluation system using machine learning for decision making, in: P. Moura Oliveira, P. Novais, L.P. Reis (Eds.), Progress in Artificial Intelligence, Springer International Publishing, Cham, 2019, pp. 27–34.
[28] K.-W. Tse, K. Hung, Behavioral biometrics scheme with keystroke and swipe dynamics for user authentication on mobile platform, in: 2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE), IEEE, 2019, pp. 125–130.
[29] A. Bal, A. Acharya, Biometric authentication and tracking system for online examination system, in: 2011 International Conference on Recent Trends in Information Systems, IEEE, 2011, pp. 209–213.
[30] G. Fenu, M. Marras, L. Boratto, A multi-biometric system for continuous student authentication in e-learning platforms, Pattern Recognit. Lett. 113 (2018) 83–92.
[31] G. Fenu, M. Marras, Controlling user access to cloud-connected mobile applications by means of biometrics, IEEE Cloud Comput. 5 (4) (2018) 47–57.
[32] M. Pleva, P. Bours, D. Hladek, J. Juhar, Using current biometrics technologies for authentication in e-learning assessment, in: 2016 International Conference on Emerging ELearning Technologies and Applications (ICETA), IEEE, 2016, pp. 269–274.
[33] R. Daza, A. Morales, J. Fierrez, R. Tolosana, MEBAL: A multimodal database for eye blink detection and attention level estimation, 2020, arXiv preprint arXiv:2006.05327.
[34] T. Soukupova, J. Cech, Eye blink detection using facial landmarks, in: 21st Computer Vision Winter Workshop, Rimske Toplice, Slovenia, 2016.
[35] A.R. Bentivoglio, S.B. Bressman, E. Cassetta, D. Carretta, P. Tonali, A. Albanese, Analysis of blink rate patterns in normal subjects, Mov. Disord. 12 (6) (1997) 1028–1034.
[36] S. Al-gawwam, M. Benaissa, Robust eye blink detection based on eye landmarks and Savitzky–Golay filtering, Information 9 (4) (2018) 93.
[37] S. Abtahi, B. Hariri, S. Shirmohammadi, Driver drowsiness monitoring based on yawning detection, in: 2011 IEEE International Instrumentation and Measurement Technology Conference, IEEE, 2011, pp. 1–4.
[38] C. Wohlin, P. Runeson, M. Höst, M.C. Ohlsson, B. Regnell, A. Wesslén, Experimentation in Software Engineering, Springer Science & Business Media, 2012.
[39] R. Woolson, Wilcoxon signed-rank test, in: Wiley Encyclopedia of Clinical Trials, Wiley Online Library, 2007, pp. 1–3.
[40] J.D. Gibbons, S. Chakraborti, Nonparametric Statistical Inference, CRC press, 2020.
[41] H.W. Lilliefors, On the Kolmogorov-Smirnov test for normality with mean and variance unknown, J. Amer. Statist. Assoc. 62 (318) (1967) 399–402.

Andrea F. Abate received the Laurea (Summa Cum Laude) in computer science from the University of Salerno, Salerno, Italy, in 1991, and a Ph.D. in applied mathematics and computer science from the University of Pisa, Pisa, Italy, in 1998. He currently serves as an Associate Professor with the University of Salerno from 2006, where he is Co-Director of the Virtual Reality Laboratory. Dr. Abate is a member of the IEEE Haptics Technical Committee and a member of the International Association for Pattern Recognition. His current research interests include multibiometric systems, virtual/augmented/mixed reality, haptics and human–computer interaction. He has authored many scientific papers published in scientific journals and proceedings of refereed international conferences and co-edited one book. He currently serves as Associate Editor for Pattern Recognition Letters and IEEE Access.

Lucia Cascone received the B.S. and M.S degrees (cum laude) in mathematics from the University of Salerno, Fisciano (SA), Italy, in 2017 and 2019, respectively. She is currently pursuing the Ph.D. degree in computer science with the Biometric and Image Processing Laboratory (BIPLAB), University of Salerno, Italy. Her research interests include applied mathematics for machine learning, biometrics, human–robot interaction and statistical analysis.

Michele Nappi received the Laurea degree (cum Laude) in computer science from the University of Salerno, Italy, in 1991, the M.Sc. degree in information and communication technology from I.I.A.S.S."E.R. Caianiello", Vietri sul Mare, Salerno, and the Ph.D. degree in applied mathematics and computer science from the University of Padova, Italy. He is currently a Full Professor of computer science with the University of Salerno. He is also a Team Leader of the Biometric and Image Processing Lab (BIPLAB). His research interests include multibiometric systems, pattern recognition, image processing, compression and indexing, multimedia databases, human–computer interaction, and VR/AR. He has coauthored more than 190 papers in international conference, peer review journals and book chapters in these fields. He was a member of IAPR. He received several international awards for scientific and research activities. He was the President of the Italian Chapter of the IEEE Biometrics Council, from 2015 to 2017.

Fabio Narducci is an assistant professor at the University of Salerno. He received a Ph.D. in Computer Science at the Virtual Reality Lab of the University of Salerno in 2015. He served as TPC of international conferences and as reviewer of several international journals. He was Guest Editor of Special Issues on IEEE Cloud Computing, Elsevier Pattern Recognition Letters and Multimedia Tools and Applications and he is currently Associate Editor of IET Electronic Letters and Editorial Board Member of Inderscience International Journal of Embedded Systems and MDPI Electronics. His research interests mainly include Machine Learning and Deep Learning, Image processing and Pattern Recognition, Biometrics, Mobile and Wearable computing.

Ignazio Passero received the Laurea degree (2005, cum laude) and the Ph.D. (2010) in computer science from the University of Salerno, Italy. He is Researcher at the University of Salerno. His research interests span the areas of semantic technologies, machine learning, embedded systems, virtual and augmented reality and gestural interfaces. He is involved in editorial and reviewer boards of several international journals and conferences and serves as reviewer for national and international scientific research projects. He is Innovation Manager in the private company Kineton (Naples, Italy).