Designing an Assistance Tool for Analyzing and Modeling Trainer Activity in Professional Training Through Simulation

FRANÇOIS ROCCA
francois.rocca@umons.ac.be
Numediart Institute – ISIA Lab, Faculty of Engineering, University of Mons
Mons, Belgium

MADISON DAVE
madison.dave@umons.ac.be
Work Psychology Unit, Faculty of Psychology and Educational Sciences, University of Mons
Mons, Belgium

VALÉRIE DUVIVIER
valerie.duvivier@umons.ac.be
Methodology and Training Unit, Faculty of Psychology and Educational Sciences, University of Mons
Mons, Belgium

AGNÈS VAN DAELE
agnes.vandaele@umons.ac.be
Work Psychology Unit, Faculty of Psychology and Educational Sciences, University of Mons
Mons, Belgium

MARC DEMEUSE
marc.demeuse@umons.ac.be
Methodology and Training Unit, Faculty of Psychology and Educational Sciences, University of Mons
Mons, Belgium

ANTOINE DEROBERTMASURE
antoine.derobertmasure@umons.ac.be
Methodology and Training Unit, Faculty of Psychology and Educational Sciences, University of Mons
Mons, Belgium

MATEI MANCAS
matei.mancas@umons.ac.be
Numediart Institute – ISIA Lab, Faculty of Engineering, University of Mons
Mons, Belgium

BERNARD GOSSELIN
bernard.gosselin@umons.ac.be
Numediart Institute – ISIA Lab, Faculty of Engineering, University of Mons
Mons, Belgium

ABSTRACT

Human audience analysis is crucial in numerous domains to understand people's behavior based on their knowledge and environment. In this paper we focus on simulation-based training which has become a popular teaching approach that requires a trainer able to manage a lot of data at the same time. We present tools that are currently being developed to help trainers from two different fields: training for practical teaching gestures and training in civil defense. In this sense, three technological blocks are built to collect and analyze data about trainers' and learners' gestures, gaze, speech and movements. The paper also discusses the future work planned for this project, including the integration of the framework into the Noldus system and its use in civil security training. Overall, the article highlights the potential of technology to improve simulation-based training and provides a roadmap for future development.

IMX '23, June 12–15, 2023, Nantes, France

ACM ISBN 979-8-4007-0028-6/23/06.
https://doi.org/10.1145/3573381.3596475

CCS CONCEPTS

• **Computing methodologies** → **Tracking**; **Activity recognition and understanding**; • **Applied computing** → **Interactive learning environments**.

KEYWORDS

Simulation-based training, Eye-tracking, Activity recognition and understanding, Assistance tool

ACM Reference Format:
FRANÇOIS ROCCA, MADISON DAVE, VALÉRIE DUVIVIER, AGNÈS VAN DAELE, MARC DEMEUSE, ANTOINE DEROBERTMASURE, MATEI MANCAS, and BERNARD GOSSELIN. 2023. Designing an Assistance Tool for Analyzing and Modeling Trainer Activity in Professional Training Through Simulation. In *ACM International Conference on Interactive Media Experiences (IMX '23), June 12–15, 2023, Nantes, France.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3573381.3596475

1 INTRODUCTION

Simulation-based training is a well-established method for training professionals in various fields, including healthcare, aviation, and defense. However, while this training method allows learners to develop their skills in situations that reproduce professional situations [16], trainers may encounter difficulties in managing it. Since simulation situations take place in a constantly evolving dynamic environment, trainers may find it challenging to consider all the information related to learners' activity that comes their way [13]. In recent years, to lighten trainers' cognitive load, advances in technology have enabled the collection of large amounts of data during training sessions, providing valuable insights into trainee performance and behavior. In this paper, we describe our efforts to develop a technological framework that integrates various components, such as microphone arrays, eye-tracking glasses, and 3D cameras, to collect and analyze data during simulation-based training which can be generalized to any audience understanding.

2 SIMULATION ENVIRONMENT AND TECHNICAL REQUIREMENTS

Even if different types of simulation training exist, all simulation training contains specific phases. In all those phases, the trainer's and learners' activity can be assisted by technology. This section presents our simulation framework and how technology is integrated into it.

2.1 Types of Simulation Training

From a global perspective, simulation represents the action of imitating, copying or reproducing [11]. This definition can be refined to refer to simulation in the field of education and training. Simulation-based training allows learners to be immersed in a realistic professional situation, directed by a scenario, and that takes place in a physical space that reproduces, to varying degrees of fidelity, the real environment of the situation [16]. Within the framework of this paper, two areas were favored to constitute the framework of analysis: training for practical teaching gestures (micro-teaching) and training in civil defense.

Micro-teaching aims to immerse learners, or future teachers, in a simulated classroom situation. During this training, future teachers are required to teach students (considered here as the audience) played by their peers. This simulated situation allows future teachers to test different teaching strategies and subsequently evaluate them. Civil defense training, on the other hand, aims to enable professionals to develop risk management skills. Professionals in the relevant fields (such as firefighters considered here as the audience) are confronted with simulated situations, including a risk to control. Even though this paper focuses on two different fields, the simulations conducted share similarities, particularly in their structures and objectives.

2.2 Simulation Phases

Therefore, the simulation situation constitutes *"the product of the learners' activity during the simulation"* [5] and allows the acquisition of professional skills through activity and reflective analysis of that activity [19]. It is traditionally divided into three phases: briefing, simulation session, and debriefing [25]. Each of these phases has its objectives, and *"trainers and learners deploy activities of different nature in these three phases"* [10]. To those three traditional phases, we have added a fourth one: the auto-debriefing phase. All those phases are presented in Figure 1.

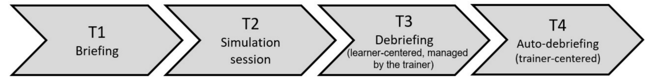

Figure 1: The different phases of the simulation.

2.2.1 Briefing (T1).
The briefing allows the trainer to present the obstacles learners will face and the pedagogical contract and establish a climate of confidence with learners. During this phase, the trainer makes a reminder of the knowledge necessary to master the situations met in the simulation.

2.2.2 Simulation Session (T2).
Simulation session constitutes the simulation itself where the audience is monitored. During this time, the trainer manages the situation and regulates the learners' activity. The trainer diagnoses the problems encountered by learners and decides to intervene if necessary.

2.2.3 Debriefing (T3).
The debriefing aims at learning through the analysis of the situation and the activity that was carried out by the learners. The trainer's mission is to encourage the learners to take a reflective look at the activity they performed during the simulation phase, to enable them to develop knowledge and skills that can be transferred to other situations.

2.2.4 Trainer-Centered Debriefing (T4).
There, the trainer takes a reflective look at their activity. It is therefore clear that the trainer's activity varies greatly when conducting training through simulation. In addition, simulation scenarios are often complex and dynamic, which makes the trainer's task even more difficult. The trainer can quickly become overwhelmed by the amount of information to process and may find themselves in a situation of cognitive overload. As a result, it complicates and questions the impact of the trainer's activity on the productive activity of the learners during the simulation (what they do during T2) and on the constructive reflection activity of the learners (what they learn from what they did during T2). In order to lighten the trainer's cognitive load, technological tools can be used to assist them.

2.3 Technological Implications

Several studies have shown that trainers, whether experienced or novice, face difficulties in observing, analyzing, and guiding learners' activity in simulation and debriefing [13, 24]. However, in order to help trainers and lighten their cognitive load, it is necessary to provide them with concrete data necessary to analyze the simulated situation. Nonetheless, since this paper focuses on two different fields, micro-teaching and civil defense, the developed tools should be adapted to all situations without disrupting the normal course of action. This also ensures that the same tools can be used to monitor any audience. Table 1 show priority information extraction according to the phase of training and field which can be divided into 3 parts : 1) extract the level of attention, 2) detect behaviors, and 3) locate speech.

Table 1: Link between technological framework, data, and time of analysis.

Micro-teaching	Civil defense
During T2, to be analyzed in T3 and T4 • Future teacher's gaze (distribution of the gaze following the zone or the students)[1] • Future teacher's movements [2] • Speech [3] **During T3, to be analyzed in T4** • Trainer's gaze [1] • Future teacher's gestures [2] • Speech [3]	**During T1, to be analyzed in T3 et T4** • Speech [3] **During T2, to be analyzed in T3 and T4** • Trainer's gaze [1] • Trainer's and learners' movements [2] • Speech [3] **During T3, to be analyzed in T4** • Trainer's gaze [1] • Trainer's and learners' gestures [2] • Speech [3]

[1] Attention Measurement, [2] Scene and Behavior Analysis, [3] Speech Localization

3 TECHNOLOGICAL FRAMEWORK AND EXPERIMENTS

The technological framework used in our project incorporates several dimensions, including attention measurement, scene and behavior analysis, and speech localization. This section provides an overview of these dimensions, including their state of the art, technology choices, and initial experimental results. The experiments were conducted in the context of micro-teaching, as the working environment was better controlled. The micro-teaching room is equipped with Noldus system [3], which records audio and video data using MediaRecorder. However, we have considered the constraints of other domains, to make our work scalable.

3.1 Attention Measurement

The human gaze does not focus on the environment in a linear manner, but rather prioritizes incoming information based on tasks or difficulty of understanding [15]. Analyzing a person's eye movement provides useful information on fatigue, concentration and task performance. It can indicate knowledge assimilation related to tasks. Eye movement analysis can also inform trainers about factors that facilitate or hinder task performance, such as the presence of distractors.

We thus decided to use the eye-tracking technology to measure attention during simulation sessions and post-simulation debriefings. By using this technology, it has been possible to collect data during simulation sessions for discussion during post-simulation debriefings. Through these sub-sections, we aim to provide a comprehensive overview of our approach to attention measurement using eye-tracking technology.

3.1.1 Attention Measurement using Eye-tracking Glasses.

We aimed to measure the visual activity of specific individuals through eye-tracking analysis to obtain a measure of attention. In micro-teaching, both the teacher's and trainer's visual activity will be studied and, in civil defense training, only the trainer's visual activity will be analyzed.

Therefore, we looked for a good ratio between eye-tracking being less invasive and having a reasonable price. After comparing three commercial solutions [14, 27, 29], we chose to focus on Pupil Labs Invisible glasses (Figure 2), which appeared to meet our expectations, and about which the research laboratory already had technical knowledge. To confirm this decision, we conducted in situ tests.

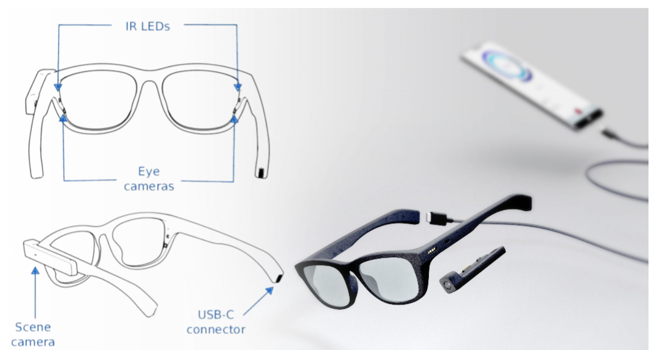

Figure 2: Pupil Labs' Invisible glasses contain cameras in their frames to track pupils and a front-facing camera to record the environment [14, 30].

As mentioned, tests were conducted in the context of micro-teaching because it allowed us to work in a more stable and controllable environment than in the civil defense training field. For this purpose, several future teachers were equipped with eye-tracking glasses, and data was recorded while they were teaching (simulation of a 30-minute teaching session on a topic freely chosen by the future teacher). We then analyzed the recordings using the enrichment tools provided with the glasses, such as face detection and QR code detection for visual attention measurement. The results were satisfying: eye-tracking works as expected and indicates with the desired precision the areas that are viewed by the wearer. Figure 3 shows the viewed area with the red circle and face detection with the blue squares. Nonetheless, it appeared that face detection was not perfect for several reasons: occlusions, suboptimal face orientation, face too far away, rapid scanning...

Regarding the QR codes, two surfaces were created: the blackboard and the background of the classroom. Figure 4 illustrate the creation of regions of interest with QR codes. Despite partial detection of the QR codes, the "background of the classroom" area

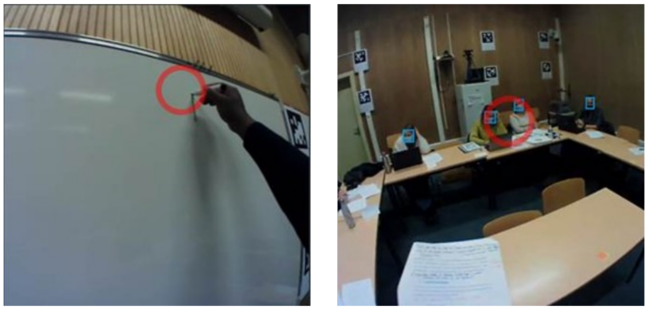

Figure 3: Gaze position (red circle) of a future teacher (here the student) in a micro-teaching environment using Pupil Invisible glasses.

remains relatively close to our definition, and heatmaps (Figure 5), which represent the areas that have been most viewed on these surfaces, could be generated.

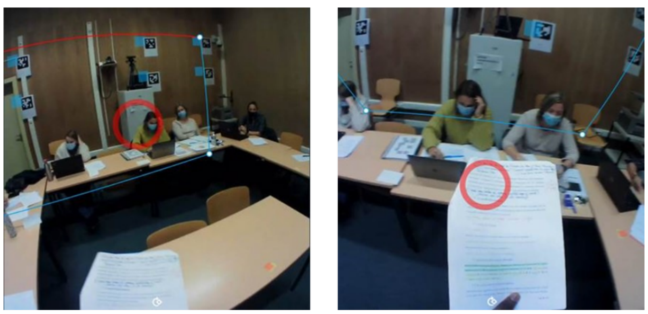

Figure 4: The region of interest, represented by a quadrilateral with blue and red edges, is determined by QR codes.

We also noticed problems with QR code detection, for similar reasons as for faces detection, but which do not seem to have a significant impact, based on the qualitative analyses carried out, which will need to be confirmed by quantitative analyses.

Other interesting eye metrics, fixations and saccades (Figure 6), could also be accessed from the tools provided by the eye-tracking glasses. Saccades are rapid eye movements that optimize information intake by positioning the attention-grabbing element on the fovea, the area of maximal resolution in the eye. Fixations, on the other hand, represent relatively stationary periods between two saccades and can be characterized by their duration and amplitude, represented by a circle.

We have tested the measures of saccades and fixations. Figure 7 illustrates the visualization of these metrics during a micro-teaching session. The results are encouraging to obtain more detailed information on visual behavior, but we will need to carefully analyze the impact of tracking errors on these metrics.

3.1.2 *Post-simulation Debriefing using Eye-tracking Data.*

Regarding the attention analysis during debriefings, a laboratory with a fixed eye-tracking system has been developed to analyze the trainers' activity during the observation and evaluation of a teaching lesson. Several possibilities were explored, such as Tobii

Figure 5: Heatmap on the area of interest defined at the back of the classroom (top) and Heatmap on the area of interest defined on the board (bottom).

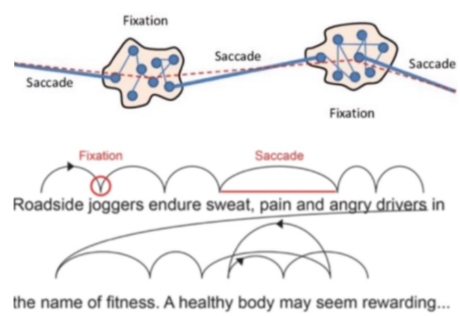

Figure 6: Examples of fixations and saccades (top: schematic diagram, bottom: fixations and saccades on text) [20]

Pro or Noldus glasses. Our choice was the GazePoint fixed eye-tracker GP3-HD, as it appears to meet our needs after testing (Figure 8) and a relatively low hardware price. Moreover, it also seems to be as reliable as more expensive systems according to several studies [2, 8].

Considering that the supervision strategies of future teachers should be extended to all actors in pre-service training [21], a comparative study of the visual strategies of future teachers (N=19) and their academic and practicum supervisors (N=10) was carried out using fixed eye-tracking technology. Each group of participants watched an excerpt (8 minutes) of an authentic teaching-learning lesson twice. On the second viewing, participants were asked to comment on the video using the Think Aloud process concurrent protocol [23]. Initial eye-tracking results indicate that, compared to future teachers, supervisors have a faster information processing

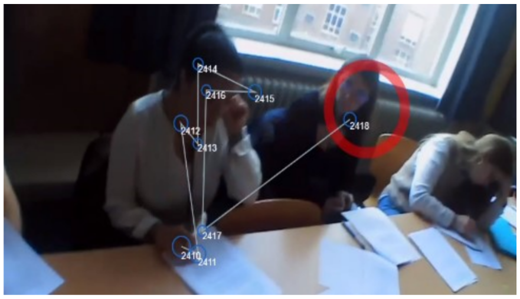

Figure 7: Saccades (in white) and fixations (in blue) of the future teacher during a simulation. The image is taken from the video showing fixations from 2410 to 2418 and shows the eye tracking path.

Figure 8: The GazePoint GP3-HD eye tracker is placed at the bottom of the screen to capture the attention of trainers during debriefing.

speed, give less fixation time to the teacher in the video and appear to spot a distracting event more slowly but have a faster return to viewing the scene.

3.2 Scene and Behavior Analysis

To analyze behavior, it is first necessary to detect and track people's movements accurately. A digital skeleton model could be used to represent an individual and easily track their movements while maintaining fidelity. The skeleton consists of several characteristic points (joints) connected by sticks (bones) and their 3D tracking over time allows for a representation of the body movements. Deep Neural Networks (DNNs) revolutionized the field of body tracking by allowing for the modeling of more complex and stable skeletons even with simple 2D cameras. OpenPose [4] is one of the pioneering algorithms in this field, and other pose detection methods, including lighter computer vision techniques, exist in platforms such as TensorFlow [1] and MMPose [6]. Once the skeleton is detected, it is tracked over time based on the proximity of the different points in the skeleton across different video frames. This approach works well in 2D but performs less effectively when one person is occluded by another. In such cases, the use of cameras capable of calculating depth provides 3D coordinates that offer greater resistance to occlusions. Cameras like the OAK-D [18] or Stereolabs ZED2 [26] allow for the extraction of depth information in ecological situations, and the ZED2 camera provides more robust person tracking and faithful movement modeling with skeleton detection methods that do not require additional commercial licenses. People tracking can be done using either skeletons or bounding boxes, providing less data than the skeleton.

For our setup, we opted for stereoscopic 3D cameras, the ZED2 cameras, which met our project's requirements in terms of resolution, refresh rate, ease of use, etc. Furthermore, the Stereolabs API already integrates OpenPose for body tracking and YoloV5 for object detection (Figure 9).

Figure 9: 3D People tracking with ZED2 camera and API (top: "bounding box" with barycenter path over time, bottom: skeleton detection) [26].

To confirm this choice, tests were conducted simultaneously with the Pupil Labs glasses during micro-teaching sessions. We recorded different scenarios with two stereoscopic cameras, placed at opposite ends of the room to cover the scene as best as possible and minimize occlusions caused by people's movements, as shown in Figure 10.

The students and the trainer were detected and had a bounding box that followed their movements. This simulation was relatively simple from the perspective of detecting and tracking people: in fact, the only person who can move freely is the teacher, with the students being relatively stationary. The results of 3D people tracking are generally satisfactory in this type of environment, although the effects of any tracking errors on the results of our measurements need to be analyzed. The next step is to synchronize the camera streams with the eye-tracking glasses.

Figure 10: Visuals of the 2 camera views with people detection and tracking, and object detection. Cameras positioned in opposition. The right parts represent the displacement of the bounding boxes.

Once people are detected and tracked in a simulation, various social signals can be measured [9, 12, 22] to gain insight into the interpersonal relationships and interactions between learners and instructors (proxemics, gestures, face expressions...).

3.3 Speech Localization

In this technological component, our primary goal is to localize speech. As our research is still ongoing, we have not selected technology yet, but working with a microphone array seems relevant.

A microphone array (Figure 11) is a system of multiple microphones arranged in a specific pattern in a space to capture sound. The network uses signal processing to analyze the differences in arrival time of sound waves at each microphone, which helps estimate the location of the sound source in three-dimensional space. Additionally, the system can enhance the desired sound while suppressing unwanted noise by analyzing the signals from each microphone in the array.

The goal is to analyze the audio in a scene to determine the orientation, and possibly the position, of sound sources, and then try to analyze emotions [17, 28] in speech and to transcribe speech to text.

4 FUTURE WORKS

We previously described the 3 technological components that will have to interact: microphone array, eye-tracking glasses, and 3D cameras. These components will provide various data during different stages of the simulation (Table 1). To ensure that the technical development is oriented towards a concrete and usable response for trainers, we have already established the intended purpose for each type of data, a representation mode for the analyses, as well as some technical remarks regarding data collection or processing. The goal

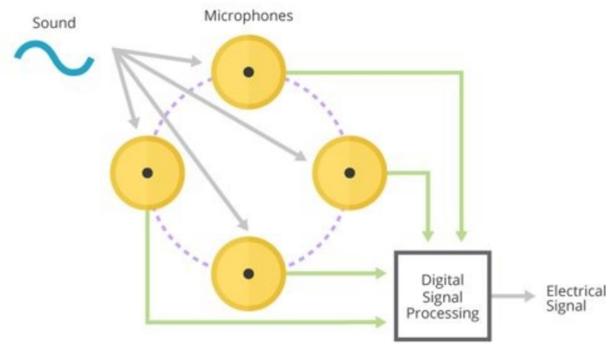

Figure 11: Schema of a sound location detection array [7].

is to work on tool development while having a clear objective for result representation (Table 2).

Once these modules are sufficiently advanced and capable of providing us with correct results, it will be possible to inject them (in whole or in part) into the Noldus system already used for observing simulation and debriefing sessions (Figure 12). We plan to perform offline analysis of the captured data and ensure that the result can be communicated and integrated as best as possible into the Noldus system.

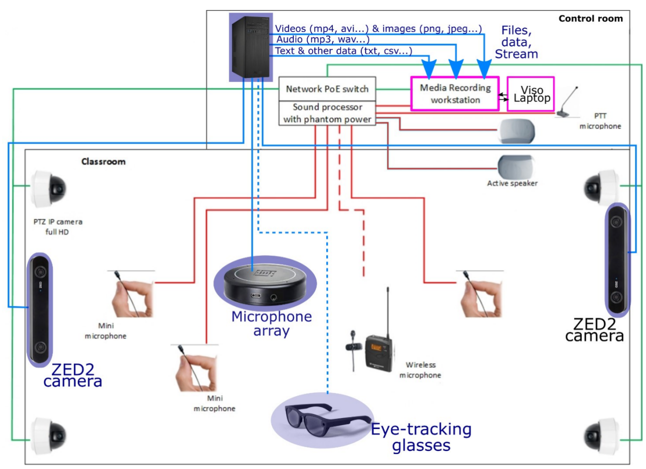

Figure 12: The figure shows the technological components that will be developed in this project (in blue) and will have to communicate with the existing system in the observation laboratory, the Noldus system for audio and video data capture.

Once the technological framework is ready, it can be used in the other planned analysis framework of this project, namely civil security training. In this regard, we are highly attentive to the technologies used and developed to ensure that they remain functional in more "wild" conditions, as we have already observed in these environments.

Table 2: The table above details, by type of collected data, the pursued purpose as well as some technical remarks regarding data acquisition or processing (in italic). This is for both domains concerned.

	Micro-teaching	Civil defense
Computer Vision (ZED2 & Eye-tracking)	**Future teacher's movements** • Heatmap of the future teacher's movements in the classroom • Heatmap for the entire duration of a micro-teaching session, as well as several maps at regular intervals › *The map should also indicate the location of the students (fixed points)* **Future teacher's gaze tracking** • Obtain a "gaze" heatmap: distribution of gaze by zone and/or by student (based on fixation times, for example) • Compare subjective view (eye-tracking glasses) to objective view (ZED2 camera) at a specific moment › *Synchronize eye-tracking camera with the ZED2* **Future teacher's pointing/gestures** • Identify moments of "break" associated with large gestures or a change in posture	**Learners' and trainer's movements** •Heatmap of the trainer's movements •Map of the movements of each trainee •Map of the movements of the group of trainees vs. the trainer **Future teacher's gaze tracking** •Obtain a "gaze" heatmap: distribution of gaze by zone and/or by student (based on fixation times, for example) •Compare subjective view (eye-tracking glasses) to objective view (ZED2 camera) at a specific moment › *Synchronize eye-tracking camera with the ZED2* **Future teacher's pointing/gestures** •Identify moments of "break" associated with large gestures or a change in posture
Audio	**Speeches (from each individual and from the future teacher vs. students)** •Identify who is speaking at a specific moment (occurrence of speeches) •Measure the speaking time of different individuals during the simulation or during a specific time period (duration of speeches) › *Synchronize the sound of the room, the sound of the future teacher's microphone,and the sound of the eye-tracking glasses*	**Speeches (from each individual and from the trainer vs. trainees)** •Identify who is speaking at a specific moment (occurrence of speeches) •Measure the speaking time of different individuals during the simulation or during a specific time period (duration of speeches) •Identify who is speaking with whom (the group of trainees vs. the trainer who is exchanging with a particular trainee)

5 CONCLUSION

In this paper, we presented a technological framework for collecting and analyzing data on any audience and we focused here on simulation-based training, using eye-tracking glasses, 3D cameras and microphone arrays. Our initial results demonstrate the potential of these tools to provide valuable insights into trainee performance and behavior, while also reducing the mental workload of trainers in audience analysis. Moving forward, we plan to continue developing these tools and integrating them into the Noldus system. We believe that this framework has the potential to enhance the effectiveness of simulation-based training in various fields, including civil security training but also, to analyze any audience watching any data content such as a performance, show, museum...

ACKNOWLEDGMENTS

This work is supported by the Project Sim'Pro - Analyser, modéliser et assister l'activité des formateurs en situation de formation professionnelle par la simulation funded by the "Fédération Wallonie-Bruxelles" through the "Action de Recherche Concertée (ARC)" Programme (ARC-21/25 UMONS5).

REFERENCES

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467* (2016).

[2] Kemeng Bai, Jianzhong Wang, Hongfeng Wang, and Xinlin Chen. 2022. A Study of Eye-Tracking Gaze Point Classification and Application Based on Conditional Random Field. *Applied Sciences* 12, 13 (2022). https://doi.org/10.3390/app12136462

[3] Noldus Information Technology BV. 2023. *MediaRecorder: Synchronous Recording of Video and Other Sources.* Retrieved March, 2023 from https://www.noldus.com/mediarecorder-human

[4] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 7291–7299.

[5] Yohann Cardin. 2016. *L'analyse de l'activité comme préalable à la conception d'un environnement virtuel de formation. Le cas d'une formation à la gestion d'incendies en milieu urbain chez les sapeurs-pompiers.* Ph. D. Dissertation. UBO. https://www.hal.inserm.fr/UNIV-UBS/tel-01278346v1

[6] MMPose Contributors. 2020. OpenMMLab Pose Estimation Toolbox and Benchmark. https://github.com/open-mmlab/mmpose.

[7] Rose B. CUI Device. 2019. *An Introduction to MEMS Microphone Arrays | CUI Devices.* Retrieved March, 2023 from https://www.cuidevices.com/blog/an-introduction-to-mems-microphone-arrays

[8] Hélio Clemente Cuve, Jelka Stojanov, Xavier Roberts-Gaal, Caroline Catmur, and Geoffrey Bird. 2022. Validation of Gazepoint low-cost eye-tracking and psychophysiology bundle. *Behavior research methods* 54, 2 (2022), 1027–1049. https://doi.org/10.3758/s13428-021-01654-x

[9] Tilman Dingler, Markus Funk, and Florian Alt. 2015. Interaction Proxemics: Combining Physical Spaces for Seamless Gesture Interaction *(PerDis '15)*. Association for Computing Machinery, New York, NY, USA, 107–114. https://doi.org/10.1145/2757710.2757722

[10] Laurie-Anna Dubois. 2017. *Apport de l'ergonomie à la formation professionnelle par la simulation: de l'analyse croisée de l'activité de formateurs, de mentors et d'aspirants-policiers à l'amélioration d'un dispositif de formation initiale.* Ph. D. Dissertation. Université de Mons (UMONS). https://hal.archives-ouvertes.fr/tel-01714061/

[11] Viviane Guéraud, Jean-Philippe Pernin, Jean-michel Cagnat, and Gloria Cortès. 1999. Environnements d'apprentissage basés sur la simulation. Outils auteur et expérimentations. *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation* 6, 1 (1999), 95–141. https://doi.org/10.3406/stice.1999.1420

[12] Edward T Hall. 1963. A system for the notation of proxemic behavior. *American anthropologist* 65, 5 (1963), 1003–1026. https://www.jstor.org/stable/668580

[13] Marc Labrucherie. 2011. Le pilotage des avions de ligne. In Ph. Alekhine & N. Pehuet (Eds.),. *Améliorer la pratique professionnelle par la simulation* (2011), 9–36.

[14] Pupil Labs. 2023. *Invisible Glasses.* Retrieved March, 2023 from https://pupil-labs.com/products/invisible/

[15] Matei Mancas, Vincent P Ferrera, Nicolas Riche, and John G Taylor. 2016. *From Human Attention to Computational Attention.* Vol. 2. Springer. https://doi.org/10.1007/978-1-4939-3435-5

[16] Allison A Murphy and Louis P Halamek. 2005. Educational Perspectives: Simulation-based training in neonatal resuscitation. *NeoReviews* 6, 11 (2005), e489–e492. https://doi.org/10.1542/neo.6-11-e489

[17] Apoorv Nandan and Jithendra Vepa. 2020. Language agnostic speech embeddings for emotion classification. In *ICML 2020 Workshop on Self-supervision in Audio and Speech.* https://slideslive.com/38930739/language-agnostic-speech-embeddings-for-emotion-classification

[18] OpenCV. 2021. *OpenCV AI Kit: OAK—D.* Retrieved March, 2023 from https://store.opencv.ai/products/oak-d

[19] Pierre Pastré. 2006. Apprendre à faire. *Apprendre et faire apprendre* (2006), 109–121.

[20] Rupert Reyneke. 2019. *Improving interactive user experience with microinteractions: An application of biometric and affect detection systems on landing pages.* Ph. D. Dissertation. Harvard University. https://www.proquest.com/openview/436114f1112f50de9be428b7c42d67f0/1?pq-origsite=gscholar&cbl=18750&diss=y

[21] Marie-Claude Rivard, Joël Beaulieu, and Mélodie Caspani. 2009. La triade : une stratégie de supervision à redéfinir! *Éducation et francophonie* 37, 1 (2009), 140–158. https://doi.org/10.7202/037657ar

[22] François Rocca, Pierre-Henri De Deken, Fabien Grisard, Matei Mancas, and Bernard Gosselin. 2015. Real-time marker-less implicit behavior tracking for user profiling in a TV context. *28th International Conference on Computer Animation and Social Agents (CASA 2015)* (2015). https://hdl.handle.net/20.500.12907/41783

[23] Katrine Roussel. 2017. Les protocoles verbaux (think-aloud protocols): enjeux méthodologiques de validité pour la recherche en contexte scolaire. *Canadian Journal for New Scholars in Education/Revue canadienne des jeunes chercheures et chercheurs en éducation* 8, 1 (2017), 160–167. https://journalhosting.ucalgary.ca/index.php/cjnse/article/view/30805

[24] Estefania Salas and Janis Cannon-Bowers. 2000. The anatomy of team training. In, S. Tobias & JD Fletcher (Eds.). *Training and retraining: A handbook for business, industry, government, and the military* (01 2000), 312–335.

[25] Renan Samurçay. 2005. *Concevoir des situations simulées pour la formation professionnelle: une approche didactique.* Toulouse : Octarès. 221–240 pages.

[26] Stereolabs. 2021. *ZED 2 - AI Stereo Camera | Stereolabs.* Retrieved March, 2023 from https://www.stereolabs.com/zed-2/

[27] Viewpoint Systems. 2023. *VPS 19 Glasses.* Retrieved March, 2023 from https://viewpointsystem.com/en/products/

[28] Noé Tits, Kevin El Haddad, and Thierry Dutoit. 2018. Asr-based features for emotion recognition: A transfer learning approach. *arXiv preprint arXiv:1805.09197* (2018). https://doi.org/10.48550/arXiv.1805.09197

[29] Tobii. 2023. *Tobii Pro Glasses 3.* Retrieved March, 2023 from https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3.

[30] Marc Tonsen, Chris Kay Baumann, and Kai Dierkes. 2020. A High-Level Description and Performance Evaluation of Pupil Invisible. *CoRR* abs/2009.00508 (2020). arXiv:2009.00508 https://arxiv.org/abs/2009.00508