Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 239 (2024) 2251–2259

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist – International Conference on Health and Social Care Information Systems and Technologies 2023

Virtual Reality on Public Speaking Phobia mitigation

Luís Ferreira[a,b,*], José Cerqueira[c], Joel Jonassi[c], Alcidio Jesus[d], Claudia Amaral[d,] Nuno Mateus-Coelho[e,f]

[a]*2Ai—Applied Artificial Intelligence Lab, School of Technology, IPCA, 4750-810 Barcelos, Portugal*
[b]*LASI – Associate Laboratory of Intelligent Systems, Guimarães, Portugal*
[c]*Polytechnic Institute of Cavado and Ave, School of Technology, 4750-810 Barcelos, Portugal*
[d]*Academia Formação do Norte, AFN, 4200-476, Porto, Portugal*
[e]*Academia Lusofona University, Copelabs, 4000, Portugal*
[f]*LAPIS2S – Laboratory of Privacy and Information Systems Security, 4000, Porto, Portugal*

Abstract

The capacity to speak in public requires personal skills that not everyone has. Anxiety and nervousness, many times easily classified as phobia, can represent serious handicaps and committing the expected speaking performance. Being a subject rather explored by contemporaneous literature, this research wants to contribute to exploring Virtual Reality and Gamification strategies for mitigating those difficulties. During this project a VR application for Google Cardboards was developed, to allow speakers to immersively face simulated 3D scenarios. With the aim of representing real-life scenarios as closely as possible, after each speaking session, the assessment process allows the user to find his limitations and reorient future experiences. Speakers are assessed on four indicators, namely voice tone, audience visual contact, head movements, and speech duration. Scenarios and speeches can be created dynamically by specialists, and classified into different difficulty levels, allowing each participant to define his own learning path. Experimentations were done with two distinct groups of students and the main results outcomes are described and analyzed. This research constitutes part of an ERASMUS+ funding project.

Keywords: Public Speaking Phobia, Virtual Reality; Unity, Gamification, Learning Path.

* Corresponding author.
E-mail address: lufer@ipca.pt

10.1016/j.procs.2024.06.416

1. Introduction

In many circumstances we find ourselves in front of a group of people, i.e., an audience, to whom we need to communicate. At a departmental dinner, an activity among friends, a conference, a class, and in many other situations, the responsibility of speaking and wanting to be understood puts some, if not a lot, of pressure on us. The ability to cope with this demand can translate into the quality of the message delivered and received. Speaking directly or in person is different from writing a news item to be read afterwards or even a message sent digitally. The real context imposes another set of values that may certainly condition.

No matter how expert we may be in an area, if we are not able to pass on that knowledge through speech or other complementary means, it will not be of much use to those who expect to benefit from that knowledge. And it is not always just a question of pedagogical limitations. *"(...) If you are effective in your speech,* behaviour, and communication, you will leave an excitement in your interlocutor. He wants to listen to you, to talk to you (...)" [1].

Public speaking several times confuses Rhetoric and Oratory. Oratory is, indeed, the art of public speaking, or “the art of speaking eloquently”. Furthermore, Rhetoric is the art of using the words effectively [2].

Many times, the fear is related with previous experiences on speech quality. Several techniques can be explored to improve speaker’s performance, such as controlling the vocalization, breathing, relaxation, tone of voice, rhythm, body movements and gestures [3]. The fear when speaking is also commonly seen as a communication disorder, any times included in the social anxiety disorders (SAD), not limited by age, gender, economic or educational situations [4]. These known occurrences made the effective public speaking (EPS) an area of continuously exploration and very interesting perspectives arose during the times [5]–[7].

This research aims to contribute to mitigating the difficulties inherent in these public speaking situations. Virtual Reality technology allows the creation of virtual scenarios that simulate real contexts. The speaker has a mechanism that helps him/her to improve or correct details that alone could not be easily overcome. A Virtual Reality mobile application was developed for use on Google Cardboards to enable participants to give speeches in virtual scenarios. Each experience is evaluated and recommendations for possible improvement are made.

2. Public Speaking Challenges

Developing public speaking skills and its inherent difficulties and attitudes like anxiety and nervousness, many times abbreviated as phobia, are subjects widely explored by literature [8]–[10]. The traditional treatments based on skills building, desensitization strategies, and cognitive restructuring are being complemented with information technology-based tools and processes [9]. From the nineties, virtual reality is being explored in therapy (VRT) in the treatment of the fear of public speaking [11], [12], and more recently, Augmented Reality (AR) and Gamification [13] start being explored as well. These evidence of studies and researches, emphasizes clearly the hard interpersonal social relations and the insufficient self-valorization and self-confidence.

In a complementary perspective, the emergence of huge sensing capacities of multiple human facets “(…) create opportunities for learners to practice while receiving feedback (…)”. The Presentation Trainer ([3] represented a great opportunity to practice a real presentation on a simulated audience. The innovative situation was the automatic feedback to overcome the incapacity to have always a human tutor to assess the speech [3], [22] . The energy used, correctness of body posture, speech rate, hard rates and many other nonverbal communication indicators were continuously presented in visual and a haptic channel (like Google Glasses) [14], [15], [19].

Context awareness is a very common and relevant requirement for social, economic, scientific, and industrial domains. Considering legacy, proactive, autonomous, and adaptive systems (and not only information systems), and human beings being an active participant in these systems, it is natural the need to continuously adapt to these dynamic environments [16] . The questions arises when the real scenario exist, and the immediate feedback is not easy to get. In the other hand, each person requires a particular strategy to better overcome emerging handicaps. The assisted and automatic feedback cannot substitute de effectiveness that only human can support [17], [20]. The public speaking performance assessment should consider different indicators for different contexts, and the way it is assessed can be determinant for the experience success [18], [21].

3. VR Training for Public Speaking

This research supported the *Third Intellectual Output (IO3) - VR Training for Public Speaking* of the ERASMUS+ funding project *Expand your Frontiers in High Technical Skills and Public Speaking*, 2020-1-FR01-KA226-VET-094712[†]. Six countries were involved, namely France, Italy, Romania, Turkey, Russia and Portugal. According to the project specification it intends an innovative training for combating social phobia of public speaking where "*(...) the Virtual Reality will introduce the participant into a simulation of public speaking and evaluate the behaviour on real time for encouraging the participant in effective public speaking and manage negative states (...)*".

The proposed platform supports a set of services for self- and guided processes of public speaking experiences. Autonomous because everyone can define their own context (real, open, or empty scenarios) and level of speech demand (open or specific topic) and guided because performance evaluation offers immediate feedback and expert people can be involved.

The platform was enriched with a set of distributed services hosted in a cloud system, that manage a public speaking sessions repository. A kind of Extended Federated Social Network (EFSN) [19], [23] prepared to support different level of participants (administrators, experts and speakers) and a set of public speaking performance indicators (duration, tone of voice, etc.).

The platform intends to improve public speaking performance and reduce negative psychological states. Virtual environments, gamification, and video game strategies over Virtual Reality (VR) technology were explored. The user is immersed in a virtual space like a conference room, a classroom, an auditorium, and others. In this space, the user complies with the task of doing a speech regarding a free theme or one proposed by the platform.

The main outputs consider:

- Supporting different scenarios in which the auditorium is empty, scenarios in which the auditorium is full, scenarios in which the auditorium is full and the audience speaking, etc. These different scenarios will represent different levels of difficulty, from simpler to more complicated.
- Automated evaluation of speech quality after collecting the speech quality indicators: i) *visual contact with the public*: the system will evaluate if the user looks at the audience (persons), and no other thins; ii) *head movement*: the system will evaluate the user's head movements to detect stress situations, etc.; iii) *voice tone*: the system will evaluate the volume level when the user speaks; and iv) *speech time*: accounting of the time the user spent on the speech.
- Immediate feedback: with the measurement of these indicators, the system generates an evaluation that is shown at the end of the activity. This evaluation is always in positive terms, to avoid undermining the user's self-confidence.

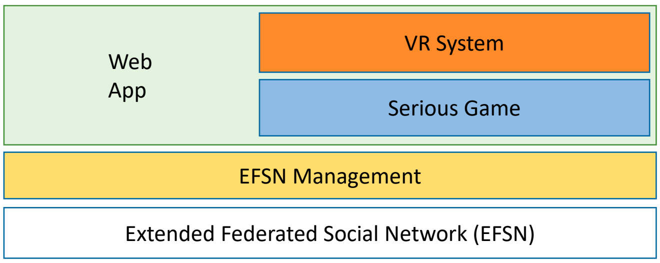


Fig. 1. EPS Framework

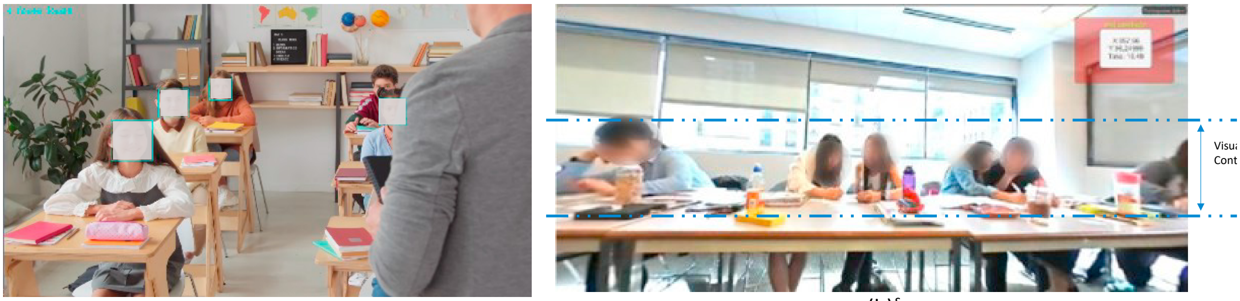
This serious VR-based game offers to the user an automated and targeted learning system, highly motivating and, at the same time, in a secure virtual environment. This application can be very useful as a practice comparing to real

[†] https://www.erasmuska2.com

situations The system is prepared to be used on Android smartphones and integrated on Google Cardboard glasses. These restrictions arrive from the initial specification of the project which demands to be widespread among the population, must be easily accessible to anyone (that's why the use of the smartphone), and have a very low cost (not more than fifteen euros). The platform also comports a set of web services to manage the expected global social network. Fig. 1 represents globally the developed framework.
The VR system (developed on Unity) integrates three logical main components: scenarios, speeches, and assessments. Each of these components offered interesting challenges to overcome a) scenarios should support 3D videos; b) speeches can be open or for specific theme and must be prepared for multi-idiom and iii) the assessment requires complex calculus over voice data.

3.1. Scenarios

The scenarios represent the simulation of the real context for the experience. To better explore all the VR potential a 3D video was suggested. The capacity to look all around could be relevant and the initial specification of 2D video and pictures would not allow that, surely. However, the visual contact assessment resulted more difficult to implement. Firstly, face detection was explored (Fig. 2 a)). Since processing algorithms require high processing capacity, the time for this task was still considerable (a few seconds) and almost certainly would not be supported by the expected devices. The second approach (Fig. 2 b)) followed a defined virtual area of visual contact (at people's head level). Anytime the users look to that region a calculus rule is activated.



(a)‡ (b)§

Fig. 2. Scenarios exploration: (a) Faces detection; (b) Visual Contact virtual layer.

3.2. Speeches

The speeches are texts that user must read. Each speech has an expected time of reading. The same speech can be in multiple idioms. Each speech has a particular level of difficulty. All these classifications are defined using JSON meta-data (Table 1) and inserted int the platform by the owner of the speech. Fig. 3 shows the moment where speeches are presented and selected by the participant.

Table 1 - JSON speech meta-data.


```
[
 {
  "FileName": "sp1_uk.txt",
  "Title":                    "Personal
Presentation",
  "Purpose": "Job application",
```


‡ *A Teacher teaching his students inside a classroom*: https://www.pexels.com/video/a-teacher-teaching-his-students-inside-a-classroom-5198159/

§ *Welcome to EC Vancouver | Learn English Classroom 360 VR*: https://www.youtube.com/watch?v=EsqQo3OuWxk

"TypeSpeech": "Training",
"Duration": 43 s,
"Dificulty": "Medium",
"Idiom": "English",
"NumberWords": 83,
"AuthorName": "Tony Kohhiba",
"DateCreation": "10/03/2023",
"DateUpdate": ""
}
]

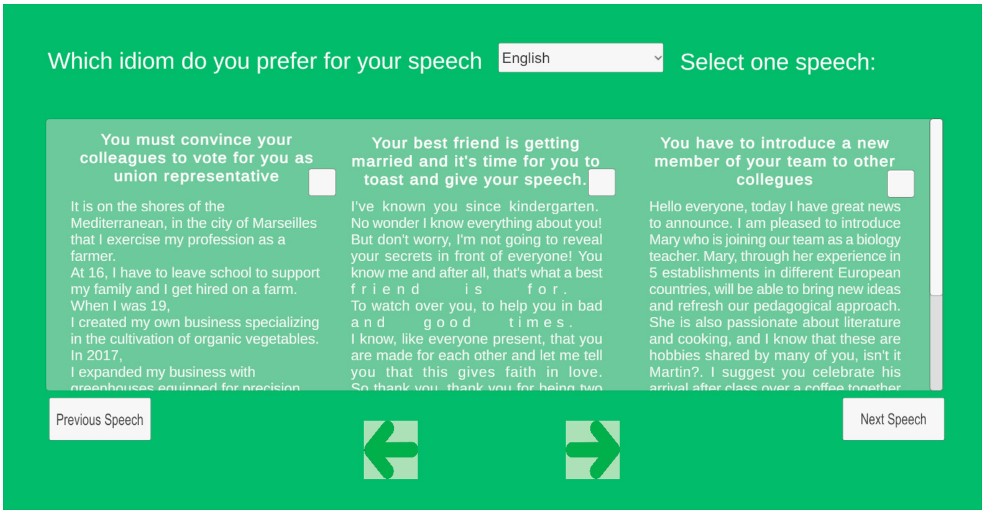


Fig. 3. Speeches manipulation.

3.3. Gamification

To promote the motivation and the spirit of resilience, some strategies of gamification were explored. Fig. 4[**] demonstrates the one to assist the speaker to where he must look at. The challenge is to align each of the circumferences with existing faces in the scenario, complementing with a kind of reward service based on the reading time, the supervised reading quality, immediate or final performance reports, etc. The level of eye contact can, in this way, be improved.

[**] *Welcome to EC Vancouver | Learn English Classroom 360 VR*: https://www.youtube.com/watch?v=EsqQo3OuWxk

Fig. 4. Gamification exploration.

3.4. Assessment

Each speech experience is analysed on four indicators: voice tone, speech duration, visual contact, and head movements.

i. the calculus of the voice tone (or tone level) produced by the human voice considers the mathematics behind sound, namely amplitude (sound´s volume), envelope (sound's amplitude over time), frequency (sound's "highness" or "lowness"), and spectrum. Each speech audio file is analysed and interpreted with *NAudio* library. Decibels level of each pitch are calculated and the number of times those values rises above a particular threshold determine the quality of the speech in terms of voice tone;
ii. each speech is defined with meta-data that includes an expected level of difficulty, duration, type, and other attributes. After each speech reading experience, the experience recorded audio´s duration is compared with the registered one;
iii. as described above, each scenario has an imaginary area that corresponds to people's head level, the *visual contact* area. Out of that area are considered no audience areas, corresponding, for instance, to the ceiling or floor of the space. The number of times those areas are looked at, represents the quality of the speech in terms of visual contact. The lesser that number, the better quality was the speech.
iv. Each speech forces the speaker to look at the audience or to avoid it. In those decisions, his head can move or not around the scenario. Each time the head moves, from one point to another, a sensor allows counting it. At the end, the lesser is that number, less nervous was the speaker and better was the speech.

The final assessment is immediately presented after speech is concluded. Figure 5[††] shows an example of such dashboard.

[††] Middle School - Math Class - 360 Video: https://www.youtube.com/watch?v=GLAo_a1i7uw

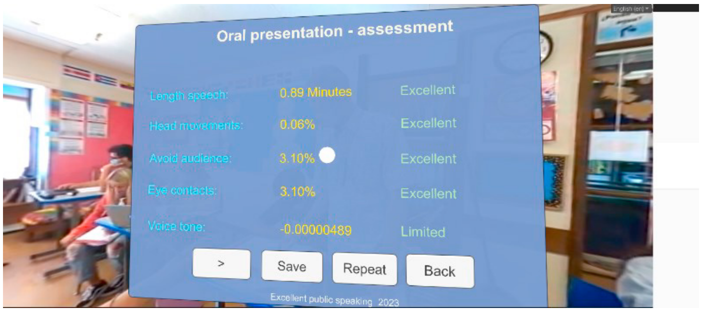


Fig. 5. Assessment dashboard.

4. Technological results

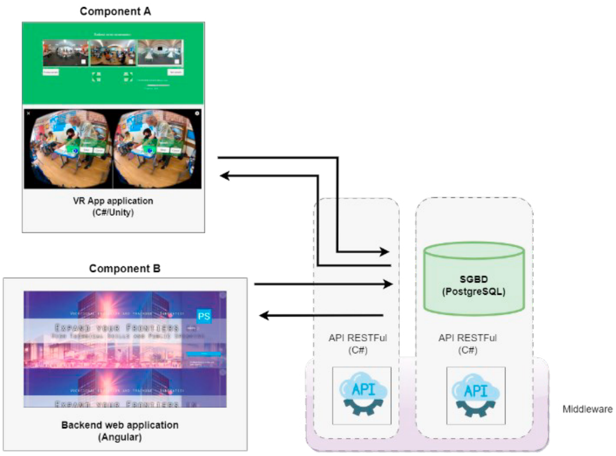


Fig. 6. VR Training Technological Architecture.

The platform represents and innovative and advanced software solution where users can behave as active participations. They can define their own path for practice and experimentation. The technological platform followed the recommended patterns of software engineering for web and mobile applications domains, in the case, *Model View Controller*. Has a system architecture (Fig. 6) with several interrelated components comporting: i) a VR App application (Component A) developed on *Unity*, that allows users to experiment with their speeches skills; ii) a Backend web application (Component B), developed on *Angular*, that supports the administration tasks (users management, moderating services, etc.), and specialists services (speeches and scenarios management; practicians monitoring, etc.); iii) a repository on top of a relational database (in PostgreSQL), to preserve all data and information; and finally iv) a middleware of RESTful web services to support all components synchronization.
The platform was developed to be prepared for future extensions, mainly services to support direct communicational channels [20]. It the future is expected the active collaboration of specialists (Psychologists, Professors, Mentors, and others) on the learning path.

5. Experimentations and Analysis of Results

The platform was explored in real scenarios. Two groups of participants of different skills were identified. One group of fifteen students from the degree in Digital Games Development Engineering and a second group of twenty senior students from the Pharmacy Technical Assistant Course of IEFP. Each participant made three speeches experiences on different type of spaces. The first group in a large auditorium the second in a classroom. The material used was Android smartphones, Google Cardboards, and Internet.

The overall results revealed curious results: i) In the beginning, almost all the participants had some difficulties in understanding well how the experiment should be carried out. The guide to the experiment explained all the necessary steps to follow, but in fact, several sessions were not carried out correctly, as many were not recorded in the database. The first group did much better; ii) after the three expected sessions, the final assessment revealed better overall performance. The chart of Fig. 7 represents the performance evolution of all the registered sessions corresponding to Group 1.

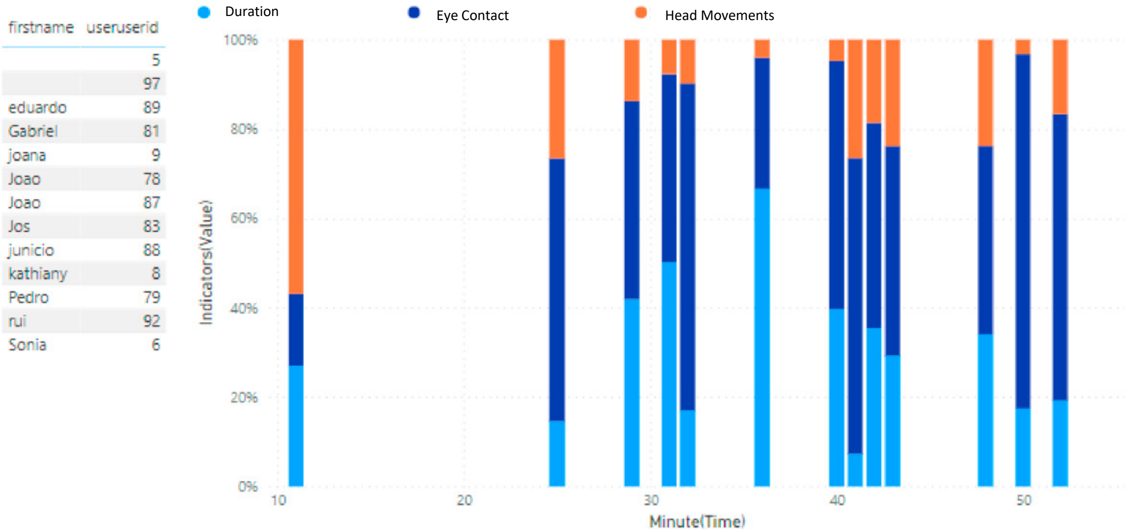


Fig. 7. Experimentation results.

After the first session (minute 11), the performance showed many head movements (57%) and few Visual Contact (15%). It can be interpreted as greater initial nervousness and some disorientation. Nevertheless, after the third (and last) session, the performance presented a much better distribution, mainly the strong decrease in nervousness, much more focus on Eye Contact, i.e., care looking at the audience, and better control of the duration of the speech. Clearly all participants mentioned that Google Cardboards conditioned a lot the experiences, as well.

6. Conclusions and Future Work

Speaking in public represents a challenge for many people, being or not experienced. The individual characteristics with other context-dependent variables, can determine the success degree of public communication. This research explored the application of Virtual Reality (VR) and Gaming strategies on mitigating such kinds of difficulties which constrain the expected performance of public speaking.

A VR application was developed to be integrated on Google Cardboards and was tested with two different groups of students. The results enlightened two important details: a) the limitations of such type of cardboards for such type of immersive interactions; and b) the relevance of practice to improve and overcome eventual speaking handicaps. All results were much better after several practice sessions.

Future works will concern two main domains: a) technologically, prepare the solutions to be used on any type of smartphone and modern VR devices; and b) expanding the platform towards an effective Extended Federated Social Network, where specialists can monitor continuously each participant and advance cognitive services, on top of artificial intelligence, will orient de better path to learn and improve each participant public speaking performance.

Acknowledgements

This paper was funded by national funds (PIDDAC), through the FCT – Fundação para a Ciência e a Tecnologia and FCT/MCTES under the scope of the projects UIDB/05549/2020 and UIDP/05549/2020

References

[1] P. R. Team, Manual Excelente Public Speaking. Expand your Frontiers in High Technical Skills and Public Speaking - 2020-1-FR01-KA226-VET-094712, 2020.

[2] A. C. Baird, "Oratory." Encyclopedia Britannica, 2023.

[3] J. Schneider, D. Börner, P. van Rosmalen, and M. Specht, "Enhancing Public Speaking Skills - An Evaluation of the Presentation Trainer in the Wild," 2016, pp. 263–276.

[4] R. C. Curtis and K. Miller, "Believing another likes or dislikes you: Behaviors making the beliefs come true.," J. Pers. Soc. Psychol., vol. 51, no. 2, pp. 284–290, Aug. 1986, doi: 10.1037/0022-3514.51.2.284.

[5] C. Baccarani and A. Bonfanti, "Effective public speaking: A conceptual framework in the corporate-communication field," Corp. Commun., vol. 20, no. 3, pp. 375–390, Aug. 2015, doi: 10.1108/CCIJ-04-2014-0025/FULL/XML.

[6] C. W. Dow, "The personality traits of effective public speakers," Q. J. Speech, vol. 27, no. 4, pp. 525–532, 2009, doi: 10.1080/00335634109380696.

[7] F. B. Robinson, Effective Public Speaking. LaSalle Extension University, 1916.

[8] A. W. Blöte, M. J. W. Kint, A. C. Miers, and P. M. Westenberg, "The relation between public speaking anxiety and social anxiety: A review," J. Anxiety Disord., vol. 23, no. 3, pp. 305–313, Apr. 2009, doi: 10.1016/J.JANXDIS.2008.11.007.

[9] E. K. Yuen et al., "A pilot of acceptance and commitment therapy for public speaking anxiety delivered with group videoconferencing and virtual reality exposure," J. Context. Behav. Sci., vol. 12, pp. 47–54, Apr. 2019, doi: 10.1016/J.JCBS.2019.01.006.

[10] Ï. Valls-Ratés, O. Niebuhr, and P. Prieto, "Unguided virtual-reality training can enhance the oral presentation skills of high-school students," Front. Commun., vol. 7, 2022, doi: 10.3389/fcomm.2022.910952.

[11] M. M. North, S. M. North, and J. R. Coble, "Virtual Reality Therapy: an Effective Treatment for the Fear of Public Speaking," Int. J. Virtual Real., vol. 3, no. 3, pp. 1–6, Jan. 1998, doi: 10.20870/IJVR.1998.3.3.2625.

[12] S. R. Harris, R. L. Kemmerling, and M. M. North, "Brief virtual reality therapy for public speaking anxiety," Cyberpsychology Behav., vol. 5, no. 6, pp. 543–550, 2002, doi: 10.1089/109493102321018187.

[13] J. M. Cerqueira, J. M. Moura, C. Sylla, and L. Ferreira, "An Augmented Reality Mathematics Serious Game," in OpenAccess Series in Informatics, 2020, vol. 81, doi: 10.4230/OASIcs.ICPEC.2020.6.

[14] C. W. Leong, L. Chen, G. Feng, C. M. Lee, and M. Mulholland, "Utilizing depth sensors for analyzing multimodal presentations: Hardware, software and toolkits," ICMI 2015 - Proc. 2015 ACM Int. Conf. Multimodal Interact., pp. 547–556, Nov. 2015, doi: 10.1145/2818346.2830605.

[15] J. Schneider, D. Börner, P. Van Rosmalen, and M. Specht, "Presentation trainer, your public speaking multimodal coach," ICMI 2015 - Proc. 2015 ACM Int. Conf. Multimodal Interact., pp. 539–546, Nov. 2015, doi: 10.1145/2818346.2830603.

[16] P. Temdee and R. Prasad, Context-Aware Communication and Computing: Applications for Smart Environment. Springer International Publishing, 2017.

[17] L. Ferreira, G. D. Putnik, N. Lopes, A. Lopes, and M. M. Cruz-Cunha, "A Cloud and Ubiquitous Architecture for Effective Environmental Sensing and Monitoring," Procedia Comput. Sci., vol. 64, pp. 1256–1262, 2015, doi: 10.1016/j.procs.2015.09.240.

[18] T. Wörtwein, L. P. Morency, M. Chollet, R. Stiefelhagen, B. Schauerte, and S. Scherer, "Multimodal public speaking performance assessment," ICMI 2015 - Proc. 2015 ACM Int. Conf. Multimodal Interact., pp. 43–50, Nov. 2015, doi: 10.1145/2818346.2820762.

[19] T. Eirinha, L. Ferreira, G. Putnik, M. C. Cunha, and N. Lopes, "Extending Federated Social Networks for effectiveness in health care," in VinOrg'15 - Fourth International Conference on Virtual and Networked Organizations Emergent Technologies and Tools, 2015.

[20] L. Ferreira et al., "A Framework for Collaborative Practices Platforms for Humans and Machines in Industry 4.0–Oriented Smart and Sustainable Manufacturing Environments," Smart Sustain. Manuf. Syst. Ind. 4.0, pp. 1–24, Jun. 2022, doi: 10.1201/9781003123866-1.

[19] Mateus-Coelho, Nuno, and Manuela Cruz-Cunha, editors.Exploring Cyber Criminals and Data Privacy Measures.IGI Global, 2023. https://doi.org/10.4018/978-1-6684-8422-7

[20] Daniel Jorge Ferreira, Nuno Mateus-Coelho and Henrique S. Mamede, Methodology for Predictive Cyber Security Risk Assessment (PCSRA), Procedia Computer Science, Volume 219, 2023, Pages 1555-1563, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2023.01.447.

[21] Mario Saraiva, Nuno Mateus-Coelho, CyberSoc Framework a Systematic Review of the State-of-Art, Procedia Computer Science, Volume 204, 2022, Pages 961-972, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2022.08.117.

[22] M. Saraiva and N. Coelho, "CyberSoc Implementation Plan,"2022 10th International Symposium on Digital Forensics and Security (ISDFS), Istanbul, Turkey, 2022, pp. 1-6, doi: 10.1109/ISDFS55398.2022.9800819.

[23] Filipe Alves, Nuno Mateus-Coelho, Manuela Cruz-Cunha, Encryption File System Framework -Proof of Concept, Procedia Computer Science, Volume 181, 2021, Pages 1237-1246, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2021.01.322.