Acceptance and Effectiveness of a Virtual Reality Public Speaking Training

Fabrizio Palmas*, Jakub E. Cichor†, David A. Plecher‡, Gudrun Klinker§
Technical University of Munich,
Garching bei München, 85748 Germany
* *Email: fabrizio.palmas@tum.de*
† *Email: jakub.cichor@tum.de*
‡ *Email: plecher@in.tum.de*
§ *Email: klinker@in.tum.de*

***Abstract*—Virtual Reality (VR) has been gaining importance due to its numerous advantages as an immersive technology for learning applications. Next to being used in fields like manufacturing, transportation, communication, retail and real estate, it has also increased in significance for human resources development. Virtual human training programs offer exposure to difficult situations without supervision in a safe and controlled environment, as they simulate nuanced interpersonal situations and therefore allow users to gain new skills and apply them to real-life situations. Furthermore, VR training engages employees in a training session by being presented with a realistic situation designed to challenge and engage them. Practising in a controlled virtual environment also allows for easy and objective measurements of user development and the identification of strengths and weaknesses. In this paper, we evaluate if there is acceptance for practising a presentation in a VR-Speech Training (VR-ST) session using six Degrees of Freedom (6DoF) and if the 44 participants of this study improved, from a subjective point of view, valuable soft skills needed to give a convincing presentation in real life. We observed a positive tendency in the acceptance and effectiveness of the VR-ST.**

***Keywords*-Virtual Reality; Virtual Training; Leadership; Social Presence; Uncanny Valley; Soft Skills; Experimental Study; Public Speaking**

I. INTRODUCTION

Immersive technologies such as Mixed, Augmented and Virtual Reality (VR) gained relevance towards the end of the 2010s. Nowadays, VR is used in various areas and in several industries. Enterprises recognize that training employees using VR offers several advantages, which leads to increased investment in VR training interventions. The purpose of a training is to prepare people for applying the newly acquired skills and attitudes in new situations and contexts. Learning transfer, conveying skills and knowledge are fundamental concepts in learning theories [1], [2]. VR enables people to feel presence and embodiment in a virtual environment [3]. It offers great potential for the development of communication skills and their transfer [4]. Employees who master these skills play a crucial role in an organization, as depending on their ability to communicate, those employees can have great impact on the company and succeed at managing the workforce. It is estimated that costs of poor communication are high for companies of all sizes [5]. Researchers have found that employees communicate mainly in meetings. They also identified the cause of their expensive costs as being correlated with inefficient coordination of human issues: poor or inadequate preparation, ineffective speakers, communication problems and too much information presented [6]. Communication in front of an audience, otherwise known as public speaking, is by most people perceived as an intimidating situation [7]. However, public speaking behavior and its associated skills can be taught, learned [8] and supported by VR training. Notwithstanding, the success and effectiveness of this kind of training depend on whether it is accepted by users and if its use is supported by companies. We investigated if there is acceptance of a VR-Speech Training (VR-ST) and its effectiveness in a sample consisting of executive MBA students.

II. THEORETICAL BACKGROUND

A. Learning in Virtual Reality

Public speaking is an important skill for clearly communicating ideas and concepts to groups of people. However, the fear of public speaking is a form of social phobia that has been reported to be one of the more common phobias in the general population [9]. It is characterized by resulting in stress responses when facing crowds during presentations. Traditional approaches to reduce person's fear employ a process called habituation, in which participants are exposed to related stressful situations repeatedly [10]. The goal of habituation is to decrease a subject's emotional response by creating conditions for getting used to the stimuli [11], [9].

In addition to interventions of this kind, VR implementations can be used with substantial effect to support existing approaches. These methods are called Virtual Reality Exposure Therapy (VRET) and work by including triggering scenarios in a VR environment such that participants can face challenges in a safe space [12]. Through these methods, VRET has been shown to elicit considerable positive effects, e.g. by providing stable results over long periods of time and showing comparable drop-out rates to traditional anxiety interventions [13]. Overall, VR interventions have produced

978-1-7281-4765-9/19/$31.00 ©2019 IEEE
DOI 10.1109/ISMAR.2019.00034

reliable decreases in self-report and physiological measures pertaining to public-speaking anxiety [14], [7], [15], [16]. VR applications in clinical contexts have also been shown to reduce the severity of agoraphobia, the fear of embarrassing situations with no possibility of escape, of acrophobia, the fear of heights, and also the fear of flying [10], [16], [17].

One way to describe the working mechanisms behind learning and training with VR applications is the flow effect. Flow describes the state of optimal experience, in which a person's skill level matches a task's challenge level, creating strong perceptions of presence and thereby placing the person in a conducive state for learning [18]. Similarly, deliberate practice has been shown to produce strong learning effects by practicing skills such that feedback is taken into account and learned skills are transferred into the next session [19].

VR can also create useful scenarios for deliberate practice, as learning trials can be repeated in the exact same manner, resulting in experimentation with new approaches and solutions. These aspects are the reason why VR is already being used in various other fields for training purposes, e.g. in engineering and surgery training [20], [21]. Similarly, serious games represent games that have been created with learning purposes in mind. Crucially, research shows that the debriefing period after playing a serious game is responsible for producing learning effects [22]. During debriefing, players receive feedback by trainers, which creates the opportunity to learn from mistakes they might have made during gameplay.

B. Virtual Reality Training

More importantly, a fundamental benefit of VR over other digital training methods is increased presence. Presence describes the concept of 'being there', inside a virtual world [23]. Presence has been identified as a primary mechanism resulting in learning effects with VR applications, as subjects have the opportunity to experience virtual situations in analogous ways as they would corresponding situations in the real world [10], [16], [24], [25]. Additionally, presence is crucial as it creates the mechanism which can make the virtual world seem as realistic as the real world, hence provoking similar reactions to the experience [24]. It has been shown that VR-based applications lead to increased perceptions of presence compared to experiencing the same applications on a desktop [26]. The approximation of the real world in VR becomes especially clear when considering that participants have been shown to react to virtual audiences in a similar way as they would to a real audience, giving credence to the idea that VR can be used for soft skills training in regard to public speaking [9], [3]. Furthermore, findings indicate that the type of audience response determines in VR how the public speaking intervention is experienced [27], [28].

As public speaking applications will necessarily have to include characters to provide sufficient simulation of the situation, social presence becomes considerably important. It is defined by the perception of feeling connected to other entities in a virtual world [23]. Social presence has been found to be significantly correlated to presence [29], to lead to increased learning outcomes in online courses [30] and to increase enjoyment in virtual experiences [31]. These findings illustrate why social presence is necessary when developing a training application.

To experience high levels of social presence, encountered characters must be accepted by participants. As the Uncanny Valley (UV) shows, acceptance cannot just be obtained by increasing a character's human-likeness and expecting simultaneous increases in familiarity [32]. The UV is a complex phenomenon, as it indicates the intricate interactions between a character's appearance and its acceptance by participants. In public speaking scenarios, characters do not have to interact in elaborate ways with participants, but instead only appear to listen and give subtle cues as feedback to the presenter. In this context, the simple addition of animations in the form of gestures have been shown to reduce related UV effects, hence increasing a character's acceptability, likability and even learning outcomes [33], [34].

Going beyond the characters however, the representation of the participant in the virtual environment is important as well. For instance, researchers have found that implementing a representation of a user's body in VR, even if the body representation is different to the user's real body, can lead to a sense of ownership for the body and appropriate reactions to possible threats [35]. These findings show how numerous elements of the VR implementation must be considered, as otherwise the experience might elicit adverse effects related to its acceptance. Nevertheless, when implementing a VR-ST, the metrics for measuring the speaking performance need to be reviewed as well.

C. Public Speaking

The following elements are important for a well received public speech:

Eye Contact: When delivering a speech, it is crucial to keep permanent eye contact with the audience by looking alternately at every auditor if there are only few people in the room. While speaking in front of a larger audience, the eye contact should alternate between different parts of the room, as it allows every auditor to feel included, hence making them more likely to pay attention. [36].

Filler Words: Filler words are words or phrases that are getting in the way of effective communication. The VR-ST identifies those words or phrases when they are used. It is very important for the audience's engagement and the speaker's confidence to avoid using filler words, as they impede the ability to pay attention. [37]

Body Language: In order to properly address the audience and increase engagement, the body, including the speaker's feet, must be facing the audience. Keeping an open hand and body position is relevant when it comes to a speech's perception. Closing the body's posture towards the audience by crossing hands or arms causes a negative evaluation. The palm should always face towards the audience, in order to make them feel addressed. Convincing gestures always happen above the waistline, as gestures underneath the waistline might not be noticed, making them seen as negative, pessimistic and restrained. The best place for gestures is the so called TV-window, an imaginary frame that lies between head and waistline. Controlled gestures to emphasize the meaning of the content are a great tool to convey a message and to make a better impression [36].

Goal: Taking all this into account, our goal was to implement a VR-ST and to use the application in a user study to explore under which circumstances a VR-ST will be accepted by participants. Moreover, we intended to investigate whether providing a report with an analysis of the participant's performance immediately after the VR-ST session would be accepted. Finally, we examined whether the VR-ST improved the participants' public speaking performance in a following presentation session.

III. Methods

A. Implementation

The VR-ST was implemented using the Unity3D game engine. We created a VR environment of a generic business meeting room that looks similar to the actual seminar room in which subjects completed their live presentation in the third phase. In the audience, we have three different characters, with the character and environmental graphics aiming to be realistic. The characters in the audience react during the VR-ST in real-time based on an audience attention system.

Tracking: For the VR-ST implementation we used the HTC Vive Head-Mounted Display (HMD), the VIVE hand controllers and two external trackers for the participants' feet. By using this hardware in combination with inverse kinematics, we are able to reconstruct the participants' gestures in human-like male or female avatar models during the VR presentation. Compared to a full motion capturing system [38], our tracking approach reduces the latency and task load. It is sufficient for measuring the required positional metrics during the virtual presentation.

Speech: In the Unity engine we get access to the integrated microphone of the HMD to get the spectrum (output) data from the microphone and to compute the decibel levels of the captured voice. We integrated the IBM Watson API to transform the participants' spoken presentations (in English) to text. The API includes hesitation markers when it discovers brief fillers or pauses in speech. Disfluency related to nonlexical utterances such as pauses and fillers can be safely filtered by hesitation markers from a transcript. The filler words the system screened for can be found in table I. In the Unity engine we analyze the transcript and identify the filler words using regular expressions. We chose this approach because it allows us to spot the filler words in the generated text with an accuracy of 100%. We counted the number of words in the data generated by the API and also calculated the time between the first word and the last word. This allows us to estimate the user's current Words per Minute (WPM). For the speech's average WPM, we take the full transcript and count the words, which we then divide by the speech's total time. The IBM Watson API provides a confidence-score as a percentage for each understood word, which states how confident the Watson API is in its predictions. A high confidence value equals a clear and understandable pronunciation.

Table I
Filler words screened by the system during presentations

like	hum (hesitations)	thing	things	see
you know	right	as I said	for instance	stuff
well	so	for example	you know what I mean	stuffs

Body Language: We generate a heat map of the body position based on the head position projected onto the floor in the room. It is calculated by obtaining the time in one position compared to the position where the user spent the maximal amount of time in the 3D environment. The eye contact with the audience is based on raycasts coming from the eyes of the HMD by intersecting the objects that participants are facing. With this method, we can measure how long participants are facing which object in the 3D environment. This approach allows us to use the gathered data to influence the audience attention system, thereby making the virtual audience react in accordance with the participant's performance. The headset rotation is based on a target spot in the meeting room, which is the center of the audience. We calculate the rotation from the headset to the center of the audience using a direction vector and a target vector to estimate the angle between the two vectors. Finally, we compare this value with threshold rotations. The threshold rotation for every value is configurable in the Unity engine. The same approach is used to gather data of the participants' viewing direction (up/forward/down).

Audience Attention: The audience reactions are based on the value "audience attention" that we generate. This value is based on the WPM (too slow or too fast), hesitations and filler words (count), body language (hand, body rotation, viewing direction), user volume (too quiet or too loud), and eye contact with the characters in the audience, balanced between the entire audience and the time spent looking at the individual characters. This metric can be interpreted as an indicator of the presentation's quality.

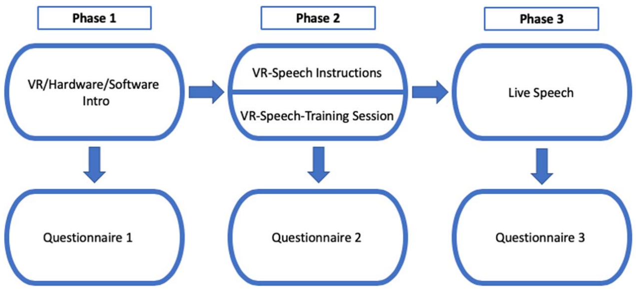

Figure 1. Chronological order of the phases of the study.

B. Study

The study was divided into three phases (figure 1): introduction, VR training session and a presentation live session. Due to the international diversity of the test group, the study was conducted in English.

Each participant completed three questionnaires during each phase of the study - the first one after the introduction, the second one immediately after the VR experience and the last one after the final live speeches. With these questionnaires, it was possible to observe the participants' acceptance for this type of training in VR and their own estimations for the improvements they made throughout the day. All participants were treated equally and tested individually.

We used questionnaires consisting of Likert scales and open-ended questions. The scaled questions were based on a 5-point scale. The respondents were required to answer the questions by indicating the extent to which they agree or disagree. For example, the participants were asked to estimate, in the first questionnaire, their own presentation skills on a scale from very bad (1) to very good (5). Another example of a scaled question was if multiple uses of the VR-ST could help the participants improve their presentation skills. Conversely, the open-ended questions aimed to gather some detailed answers and feedback.

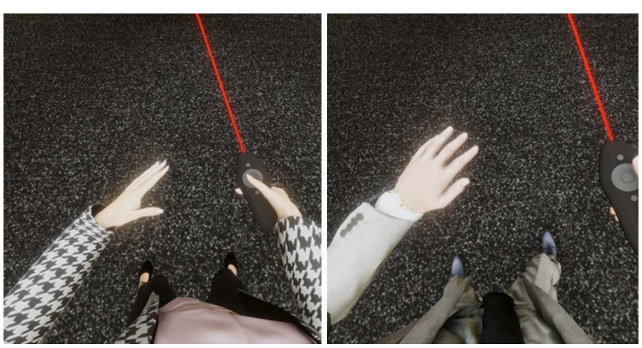

Figure 2. Embodiment in VR: female (left) and male (right) avatars seen from the point of view of the participants.

Phase 1: All the participants received an explanation of the functionality of the VR-ST and its hardware during this phase. Afterwards, all participants prepared their own presentation for the VR-ST, which was the same presentation used by the participants in the third phase for the final live session. All the presentation were held in English.

Phase 2: How the VR-ST is supposed to be used was explained separately to every participant just before starting the VR-ST session. Once the avatar is chosen in the VR environment, each participant estimates the difficulty for each slide of their presentation. The difficulty levels low, medium and high are correlated with the amount of time the participants want to invest in each slide. This step is relevant to help the participants improve their time management skills, while also balancing the attention system of single characters in the audience during the VR-ST session.

By clicking on the ”Go to Office” button, the setting changes and the participants find themselves in a virtual office with three people sitting around a table, with a laptop being on the table in front of the participant. The time counter starts by clicking on the ”Start” button on the virtual computer screen. The participants are then able to see their own presentation on the laptop or behind them on the screen. The total time for the VR-ST without the setup phase was 5 minutes per participant. The speech is finished by clicking on ”Stop”. All participants received a detailed analysis documented in a Portable Document Format (PDF) report with their personal feedback, immediately after the VR-ST session by email. Prior research has hypothesized that one major advantage of VR training applications would be the ability of giving feedback in innovative and personalized ways [4]. This feedback is generated by the individual participant's gathered data during the VR-ST session.

Phase 3: In this phase, the participants give their own presentation used in phase 2 in a real-life situation in front of a real audience. Crucially, the meeting room and the speech situation look similar to the VR environment.

C. Metrics

The measures were evaluated fully automatically by the VR-ST and are listed in table II. Those are shown to the participants in their report at the end of phase 2. These metrics are also used by the adaptive attention system to balance the audience's engagement and their reactions in real time (figure 3). During a VR-ST, the participants should analyze performance and adapt their public speaking behaviour by observing the audience's engagement and reacting accordingly. This process improves their presentation and communication skills by regaining their attention and searching for eye contact, adapting the speed of the spoken words, changing posture, gestures or position in the space. Signs of disinterest in the audience are the following: discontinued eye contact, dishonest smiles, heads resting on hands, blocking (crossed arms and clenched fists), pacifiers (cleaning a spot on the table), looking at the phone or at the watch and acting nervously through hand movements.

Table II
METRICS USED FOR THE ADAPTIVE ATTENTION SYSTEM FOR THE AUDIENCE AND FOR GENERATING THE REPORT FOR EACH PARTICIPANT

Metric	Variables	Optimal Level
Timing	Self-estimation of the presentation's time and actual presentation time	Elapsed time = estimated time = max. time
Transcript	Speech transcription (accuracy)	Accurate transcript, no filler words
Eye Contact	Viewed objects in virtual environment	Audience exclusively
Voice	Speaking volume	Balanced
Positioning	Subject position in virtual environment	Subject moves around in the virtual environment
Filler Words	Hesitations and fillers	No filler words
Body Language	Body facing direction, hand position	Facing audience, hand position in TV-window

Figure 3. Comparison between engaged (left) and bored (right) virtual audience.

The report for the participants provides feedback, explaining how to interpret the following metrics:

1) Timing:
The timing metric consists of three distinct features: elapsed time, estimated time and maximum time. Elapsed time represents the actual duration of the VR presentation, while estimated time shows an estimation of the time required for the presentation based on the complexity of the slides. With maximum time, we show the participants a given or chosen time-limit.
2) Transcript:
The transcript of the spoken presentation based on what is recognized by voice recognition. Filler words are written in red.
3) Eye contact:
We track the objects which are viewed by the participants. In the report, a pie chart shows which objects were viewed during the presentation. The chart is divided in audience, laptop, screen, floor, and others. With a heatmap, we show from two different points of view where the subjects looked at exactly during the presentation (figure 4). Frequent points of view are depicted on a scale between green and red, while rare ones are represented in blue.

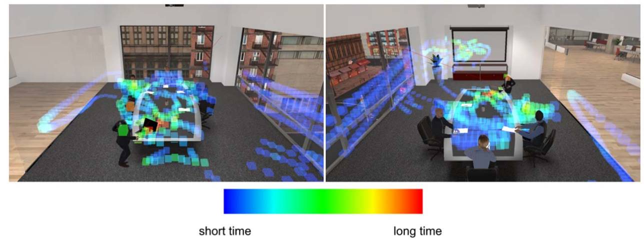

Figure 4. Heatmap based on what the participants look at.

4) Voice:
Regarding the voice, we attempted to measure its quality on three dimensions: volume, confidence and WPM. The volume is displayed as a graph showing its changes throughout the presentation, as well as the average voice volume. An average around zero equals a loud and understandable voice, whereas values below and above zero indicate too quiet or too loud voice projection, respectively. Confidence is based on the confidence-score and the correlated pronunciation quality for spoken words. Additionally, a graph shows the spoken WPM (figure 5). An average between 100 and 150 words per minute is regarded as a pleasant speaking speed.

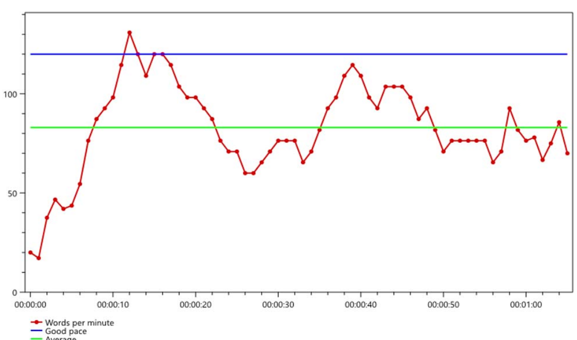

Figure 5. Example of a graphic with words per minute in the report.

5) Positioning:
A heatmap of where the participants were standing during their presentation and whether they moved (figure 6). Frequent points are depicted green to red, rare ones in blue.
6) Filler words:
A listing of filler words said during the presentation and the amount of times they were used (figure 7). In the transcript they are written in red.
7) Body language:
We provide information to the participants regarding their body language through metrics pertaining to viewing direction, hand position and body direc-

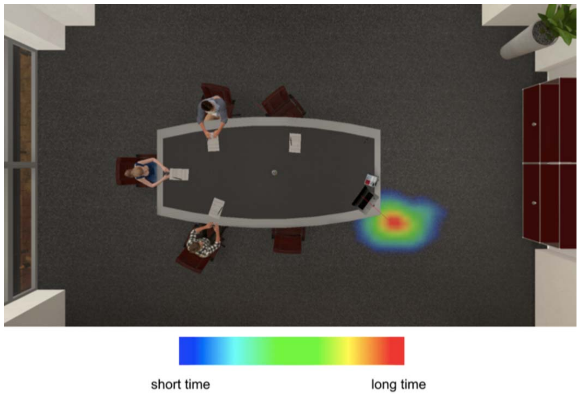

Figure 6. Heatmap based on the physical movement of the participant.

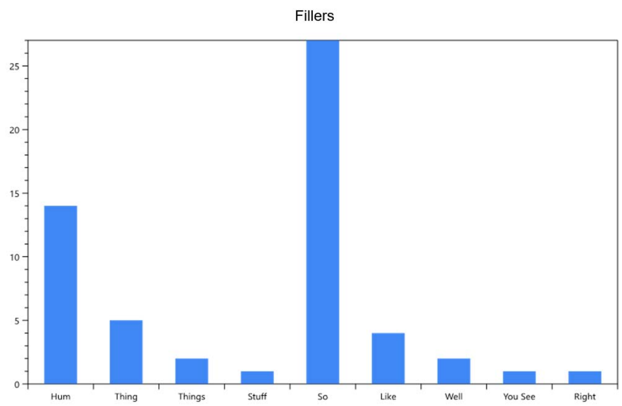

Figure 7. Example of a graphic with filler words in the report.

tion (figure 8). These metrics are described through pie charts, where the viewing direction chart shows how often the speaker looked forwards, upwards or downwards. Regarding the hand position, a pie chart displays how frequently subjects held their hands at a medium, high or low height. The final pie chart reveals how often the participant's body faced forwards, sideways or backwards to the audience.

8) Response analysis:
A graph showing the attention of the virtual audience during the presentation (figure 9). The attention is influenced by eye contact, voice, gestures and body language.

IV. Results

The following collected data resulted from the answers given by each person at the end of each phase during this study. The sample consisted of 44 participants (15 female and 29 male) students of an executive Master of Business Administration (MBA). Their ages ranged from 24 to 50 and the average age was 34.8. Every participant was employed at the time of this study and the average

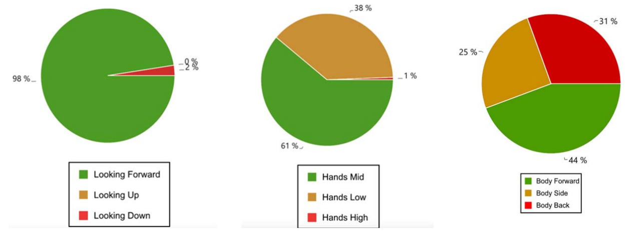

Figure 8. This graphic of the report gives some indications for performance based on the body language.

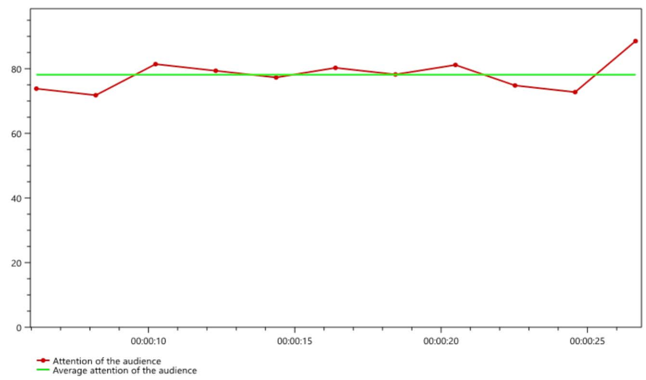

Figure 9. Attention of the virtual audience during the presentation. The attention is influenced by the eye contact, voice, gestures and body language.

working experience in the actual company was 3.4 years. In addition, the respondents' average working experience was 10.7 years. 25 persons participated at least once in a presentation course or training prior to this study, while the other 19 participants did not. Of the 15 female participants, 8 had a previous course or training about presenting, while the remaining 7 had no previous experience. Five of the 8 above rarely used VR whereas 3 of the remaining never used VR before. Half of the participants have used VR rarely, while the other half never used it before. All subjects except one wanted to improve at least one of the following presentation skills: body language, audience engagement, time management, voice or confidence.

The overall mean value of the WPM was 138.18 and the average value of the filler words was 33.16. Table III shows a high standard deviation for the WPM and the filler words. The Pearson correlation coefficient R of 0.1089 shows that the relationship between the WPM and the occurrence of filler words is weak.

Female participants who did a previous course or training on presentation skills at some point in their life, held eye contact with the audience 57.87% of the time, while female subjects who did not complete a course previously only looked towards the audience 35.71% of the time. The percentage of eye contact with the audience for female and male participants that used VR rarely was higher compared

Table III
FILLERWORDS

	WPM	Filler Words
Sum	-	1459
Mean	138.18	33.16
Variance	469.87	785.95
Standard Deviation	21.68	28.03

Table IV
EYE CONTACT

Eye contact with the audience		
Use of VR	**Never(%)**	**Rarely(%)**
Male	37.85	54.78
Female	51.77	57.87

to the group of subjects who never used VR before (table IV).

Even though the participants were able to move freely in the real environment, they mostly chose to remain static in the spawn position. The mean values in table V show that the participants had high compliance for a correct attitude during the speech and they were able to achieve an average of 80.22% for audience attention.

Table V
BODY LANGUAGE

	Mean(%)
Hands position (middle)	81.90
Body Direction (forward)	80.00
Looking Direction (forward)	95.90
Confidence	76.79
Audience Attention	80.22

The participants' self-evaluations showed that the use of VR-ST was able to help the majority to overcome the fear of public speaking: 46% mostly agreed and 18% completely agreed. Additionally, 80% of the participants indicated that according to their perception, multiple sessions of the VR-ST would likely help in improving their presentation skills: 41% affirmed ”very much” and 39% ”much”. 92% is the relative frequency of participants that had ”fun” using the VR-ST. 72% stated it with 5 points and 21% with 4 points. The results of the third questionnaire show that 81% of the participants considered the feedback on the report helpful to improve their presentation skills. 34% of the group found it ”very helpful” and 47% ”helpful”. The group also found the VR technology and the use of it in training ”very useful”. The median and the mode of these answers is 5 points. 42 of 44 participants affirm to have improved at least one of their presentation skills. At least one of the following skills was marked as improved: voice, body language, audience, engagement, confidence and/or time management.

V. DISCUSSION AND LIMITATIONS

The self-reported effectiveness measures provide a first data set with enough information to investigate whether the participants accepted a training of this kind. The results showed a tendency that the VR-ST was accepted and effective. This acknowledgement gives us a solid base for future studies with the goal of measuring the objective effectiveness of the VR-ST. The acceptance rate of all measured parameters was positive for the majority of the participants. The answers of the open-ended questions show that the subjects found it challenging to get used to the technology and to the usage of the VR-ST in such a short amount of time. 95% of the participants stated at the end of the third phase to have improved their presentation skills. That means that just 5% of the participants felt that they did not improve any skills. From the beginning of the first session, there was just one person who affirmed that he or she did not want to improve any presentation skills. Moreover, the fact that female participants who attended a presentation course at some point in their life, spent more time looking towards the audience, shows that presentation skills obtained in the real world seem to translate to the virtual world as well. This finding supports the point that learning and transfer can be established between the virtual and real world. However, our sample size of 44 participants was comparatively small and should be extended in future studies.

Nevertheless, one single short training session did not suffice to objectively evaluate improvements in a real-life public speaking session. With many participants stating that multiple sessions could help them to further improve their skills, multiple training sessions would be needed to achieve better training results and effectiveness. The report was accepted and it seems that its usage in combination with the VR-ST supports learning effects. The majority of the participants stated that VR-ST could also help them to overcome the fear of public speaking. This confirms that VR is identified by this group as a safe environment. 80% of the subjects stated the VR-ST could be a useful training method for companies. A VR-ST has the potential of being considered fun, especially when it is challenging enough to match the participants' skill level, thereby activating the flow effect. However, some participants stated that it could not be implemented in their own company ”because of conservative decision makers and skepticism” or the company being ”too small, too conservative”. This is a clear sign that the real challenge to establish this kind of VR training will also be its acceptance by companies. Organizations need to embrace and understand the potential of this technology, which would make its widespread application more likely. By accepting to use the VR-ST, employees have the possibility to improve their skills and communicate in a more efficient way that could lead to increased efficiency and productivity for companies and individuals alike.

VI. Future Work

The main idea for implementations in the future would be to extend the control group. The transfer of acquired skills should be verified through impartial methods. One possibility will be adding a second VR-ST session to help participants with staying focused on their performance during the VR-ST session. This additional session would provide significant advantages over the prior implementation by allowing for more time to get used to the technology. Moreover, it could reduce the likelihood of participants new to the technology being afraid of making mistakes by using VR in an unintended manner. We are currently planning a long-term intervention study as the next step, which will give us the opportunity to use our metrics directly as objective measures and compare the participants' results between measuring points over a span of multiple weeks. We also plan to develop a gamified version of the VR-ST to test if we can transfer skills in a more efficient and engaging way. This implementation will also have to be tested for efficiency and acceptance by participants.

VII. Conclusion

The acceptance of the VR-ST is one of the most important factors that ensure that new skills can be taught using this technology. VR-ST could support the learning process and improve the communication skills in public speaking if accepted by its users. Furthermore, the acceptance of VR is not just necessary at an individual level but also at the company level. If enterprises want to take advantage of its full potential and allow employees to improve their performance, they should embrace the use of this technology for training. As our results demonstrate, a group of executive MBA students accepted the use of the VR technology and of the VR-ST, which led to positive learning outcomes. The effectiveness of the VR-ST was confirmed from their personal perspective.

References

[1] A. Marini and R. Gnreux, "The challenge of teaching for transfer." in *Teaching for transfer: Fostering generalization in learning.* Mahwah, N. J.: Erlbaum, 1995, pp. 1–19.

[2] N. R. Pennington, N. and J. Rahm, "Transfer of training between cognitive subskills: Is knowledge use specific?." *Cognitive Psychology*, vol. 28, no. 2, pp. 175–224, 1995.

[3] M. Slater, C. Guger, G. Edlinger, R. Leeb, G. Pfurtscheller, A. Antley, M. Garau, A. Brogni, and D. Friedman, "Analysis of physiological responses to a social situation in an immersive virtual environment," *Presence: Teleoperators and Virtual Environments*, vol. 15, no. 5, pp. 553–569, 2006.

[4] M. Schmid Mast, E. P. Kleinlogel, B. Tur, and M. Bachmann, "The future of interpersonal skills development: Immersive virtual reality training with virtual humans," *Human Resource Development Quarterly*, vol. 29, no. 2, pp. 125–141, 2018.

[5] P. M. Buhler and J. D. Worden, *Up, Down, and Sideways: High-impact Verbal Communication for HR Professionals.* Society for Human Resource Management, 2013.

[6] N. C. Romano and J. F. Nunamaker, "Meeting analysis: Findings from research and practice," in *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*. IEEE, 2001, pp. 13–pp.

[7] S. R. Harris, R. L. Kemmerling, and M. M. North, "Brief virtual reality therapy for public speaking anxiety," *Cyberpsychology & behavior*, vol. 5, no. 6, pp. 543–550, 2002.

[8] S. B. Fawcett and L. K. Miller, "Training publicspeaking behavior: An experimental analysis and social validation 1." *Journal of Applied Behavior Analysis*, vol. 8, no. 2, pp. 125–135, 1975.

[9] D.-P. Pertaub, M. Slater, and C. Barker, "An experiment on public speaking anxiety in response to three different types of virtual audience," *Presence: Teleoperators & Virtual Environments*, vol. 11, no. 1, pp. 68–78, 2002.

[10] L. F. Hodges, R. Kooper, T. C. Meyer, B. O. Rothbaum, D. Opdyke, J. J. De Graaff, J. S. Williford, and M. M. North, "Virtual environments for treating the fear of heights," *Computer*, vol. 28, no. 7, pp. 27–34, 1995.

[11] I. B. Mauss, F. H. Wilhelm, and J. J. Gross, "Autonomic recovery and habituation in social anxiety," *Psychophysiology*, vol. 40, no. 4, pp. 648–653, 2003.

[12] K. Meyerbröker and P. M. Emmelkamp, "Virtual reality exposure therapy in anxiety disorders: a systematic review of process-and-outcome studies," *Depression and anxiety*, vol. 27, no. 10, pp. 933–944, 2010.

[13] D. Opriş, S. Pintea, A. García-Palacios, C. Botella, Ş. Szamosközi, and D. David, "Virtual reality exposure therapy in anxiety disorders: a quantitative meta-analysis," *Depression and anxiety*, vol. 29, no. 2, pp. 85–93, 2012.

[14] C. A. Anderson, A. Shibuya, N. Ihori, E. L. Swing, B. J. Bushman, A. Sakamoto, H. R. Rothstein, and M. Saleem, "Violent video game effects on aggression, empathy, and prosocial behavior in eastern and western countries: A meta-analytic review." *Psychological bulletin*, vol. 136, no. 2, p. 151, 2010.

[15] M. D. Nazligul, M. Yilmaz, U. Gulec, M. A. Gozcu, R. V. OConnor, and P. M. Clarke, "Overcoming public speaking anxiety of software engineers using virtual reality exposure therapy," in *European Conference on Software Process Improvement*. Springer, 2017, pp. 191–202.

[16] M. M. North, S. M. North, and J. R. Coble, "Virtual reality therapy: An effective treatment for psychological," *Virtual reality in neuro-psycho-physiology: Cognitive, clinical and methodological issues in assessment and rehabilitation*, vol. 44, p. 59, 1997.

[17] B. O. Rothbaum, L. Hodges, S. Smith, J. H. Lee, and L. Price, "A controlled study of virtual reality exposure therapy for the fear of flying." *Journal of consulting and Clinical Psychology*, vol. 68, no. 6, p. 1020, 2000.

[18] J. Nakamura and M. Csikszentmihalyi, "The concept of flow," in *Flow and the foundations of positive psychology*. Springer, 2014, pp. 239–263.

[19] K. A. Ericsson and N. Charness, "Expert performance: Its structure and acquisition." *American psychologist*, vol. 49, no. 8, p. 725, 1994.

[20] A.-H. G. Abulrub, A. N. Attridge, and M. A. Williams, "Virtual reality in engineering education: The future of creative learning," in *Global Engineering Education Conference (EDUCON), 2011 IEEE*. IEEE, 2011, pp. 751–757.

[21] S. Haque and S. Srinivasan, "A meta-analysis of the training effectiveness of virtual reality surgical simulators," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, pp. 51–58, 2006.

[22] R. Garris, R. Ahlers, and J. E. Driskell, "Games, motivation, and learning: A research and practice model," *Simulation & gaming*, vol. 33, no. 4, pp. 441–467, 2002.

[23] S. T. Bulu, "Place presence, social presence, co-presence, and satisfaction in virtual worlds," *Computers & Education*, vol. 58, no. 1, pp. 154–161, 2012.

[24] M. V. Sanchez-Vives and M. Slater, "From presence to consciousness through virtual reality," *Nature Reviews Neuroscience*, vol. 6, no. 4, p. 332, 2005.

[25] M. Slater, D.-P. Pertaub, C. Barker, and D. M. Clark, "An experimental study on fear of public speaking using a virtual environment," *CyberPsychology & Behavior*, vol. 9, no. 5, pp. 627–633, 2006.

[26] R. Moreno and R. E. Mayer, "Learning science in virtual reality multimedia environments: Role of methods and media." *Journal of educational psychology*, vol. 94, no. 3, p. 598, 2002.

[33] S. W. Cook, H. S. Friedman, K. A. Duggan, J. Cui, and V. Popescu, "Hand gesture and mathematics learning: lessons from an avatar," *Cognitive science*, vol. 41, no. 2, pp. 518–535, 2017.

[27] D. Pertaub, M. Slater, and C. Barker, "An experiment on fear of public speaking in virtual reality," *Studies in health technology and informatics*, pp. 372–378, 2001.

[28] M. Slater, D.-P. Pertaub, and A. Steed, "Public speaking in virtual reality: Facing an audience of avatars," *IEEE Computer Graphics and Applications*, vol. 19, no. 2, pp. 6–9, 1999.

[29] S. Thie and J. van Wijk, "A general theory on presence," *1st Int. Wkshp. on Presence*, 1998.

[30] J. Richardson and K. Swan, "Examing social presence in online courses in relation to students' perceived learning and satisfaction," *JALN*, 2003.

[31] C. S. Oh, J. N. Bailenson, and G. F. Welch, "A systematic review of social presence: Definition, antecedents, and implications," *Front. Robot. AI 5: 114. doi: 10.3389/frobt*, 2018.

[32] M. Mori, "The uncanny valley," *Energy*, vol. 7, no. 4, pp. 33–35, 1970.

[34] L. Piwek, L. S. McKay, and F. E. Pollick, "Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion," *Cognition*, vol. 130, no. 3, pp. 271–277, 2014.

[35] M. Slater, B. Spanlang, M. V. Sanchez-Vives, and O. Blanke, "First person experience of body transfer in virtual reality," *PloS one*, vol. 5, no. 5, p. e10564, 2010.

[36] J. Baier, *Körpersprache. Was sie mit Körpersprache bewirken können ein Trainingsbuch*. Gabal Verlage, 2016.

[37] H. H. Clark and T. Wasow, "Repeating words in spontaneous speech." *Cognitive psychology*, vol. 37, no. 3, pp. 201–242, 1998.

[38] D. Roth, J.-L. Lugrin, J. Büser, G. Bente, A. Fuhrmann, and M. E. Latoschik, "A simplified inverse kinematic approach for embodied vr applications," in *2016 IEEE Virtual Reality (VR)*. IEEE, 2016, pp. 275–276.