VR Public Speaking Simulations Can Make Voices Stronger and More Effortful

Ïo Valls-Ratés[1,3][a], Oliver Niebuhr[1][b] and Pilar Prieto[2,3][c]
[1]*University of Southern Denmark, Alsion, 2, Sønderborg, Denmark*
[2]*Universitat Pompeu Fabra, Roc Boronat, 138, Barcelona, Spain*
[3]*Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig de Lluís Companys, 23, Barcelona, Spain*

Keywords: Virtual Reality, Public Speaking, Prosody, Voice.

Abstract: In the field of public speaking, studies have mainly centered on the effects of virtual reality (VR) environments in reducing public speaking anxiety (PSA). However, prior research on the effect of VR simulations on high-school students' performance in terms of the prosody of their speech and number of gestures while being immersed in a VR scenario is limited to just one study. The present paper examines the effects of practicing speeches with a VR-simulated audience on self-perceived PSA, and speaking performance qualified on the basis of the prosodic characteristics of the presenter's voice and the rate of gestures they use while presenting. Forty-seven high school students participated in either a VR group that practiced a two-minute speech in front of a virtual audience, or a Non-VR group that delivered the same oral presentation alone in a room. Crucially, these were compared with a baseline initial oral task where students presented in front of a live audience. Practicing with VR resulted in significant differences across the groups pointing to VR-trained voices becoming stronger, more effortful and louder. Simulated audience seems to help speakers develop more audience-oriented prosody. This is particularly useful for rehearsing public speaking skills in the context of secondary school education to improve students' oral competence.

1 INTRODUCTION

Practicing public speaking in classrooms is crucial for students' confidence and social development (King, 2002; Iberri-Shea, 2009). It enhances skills such as decision-making, critical thinking, and empathy (Schneider et al., 2017). This skillset is vital for future professional success (Nguyen, 2015) and can alleviate public speaking anxiety (Liao, 2014). Educational institutions should recognize its importance in fostering self-confidence and self-directed learning (Munby, 2011), while instructors play a key role in motivating student engagement (Kaufmann & Tatum, 2017).

2 RELATED WORK

Against this background, the aim of this study was to investigate the effects of VR on adolescents' voice and gesture, and their self-perceived anxiety while being immersed in a VR setting performing a speech.

Public-speaking training is rarely offered in educational settings. This is unfortunate, given that, according to Ford and Wolvin (1993), once public speaking is trained in the classroom, students perceive that their communication becomes more effective and they feel more self-confident, more confident that they are well-regarded by others, more able to reason with other people and more skillful at using language appropriately. The ability to communicate effectively and appropriately is learned and, therefore, must be taught (Morreale et al., 2000). In short, teaching public speaking skills needs to be directly integrated into the classroom.

Mastering public speaking shares notable parallels with the process of acquiring a new language. Research suggests a critical time window exists for language acquisition, as demonstrated by studies (e.g.,

[a] https://orcid.org/0000-0001-9511-6927
[b] https://orcid.org/0000-0002-8623-1680
[c] https://orcid.org/0000-0001-8175-1081

Valls-Ratés, Ï., Niebuhr, O. and Prieto, P.
VR Public Speaking Simulations Can Make Voices Stronger and More Effortful.
DOI: 10.5220/0012748900003693

In *Proceedings of the 16th International Conference on Computer Supported Education (CSEDU 2024) - Volume 1*, pages 685-693
ISBN: 978-989-758-697-2; ISSN: 2184-5026

Hartshorne et al., 2018). This critical period, shown to extend to aspects such as prosody—the rhythmic, intonational, and stress patterns defining a language (Huang & Jun, 2011)—underscores the importance of timely exposure and practice in developing linguistic fluency. Just as proficiency in a new language benefits from early immersion, effective public speaking skills thrive with deliberate training and experience, emphasizing the need for strategic interventions to optimize learning outcomes, especially in the classrooms.

A technology that can help engaging students in the learning process is virtual-reality (henceforth VR). VR technology displays three-dimensional computer-generated scenes which create the illusion for the viewer that he or she is physically located within that simulated space and interacting with it, in other words, a sensation of physical presence (Radianti et al., 2020). VR has been used in a wide variety of fields to treat phobias and post traumatic disorder (Baños et al., 2011), in the entertainment industry (Bianchi-Berthouze, 2013) and for medical rehabilitation purposes (Bourdin et al., 2019). Several studies have also assessed VR as a tool to be included in speech communication courses to enhance students' performance and make learning more meaningful, and to reduce public-speaking anxiety (henceforth PSA). For instance, in a meta-analysis exploring PSA and VR, Hui Lim et al. (2022) identified a total of 92 studies and they concluded that since the results of using VR were similar to those obtained by other modes of therapy, it can be considered an effective tool to treat clinical PSA (see also Daniels et al., 2020) and appropriate as a complementary method to other therapies such as cognitive behavior therapy. Indeed, other studies have shown how combining VR with other modalities of therapy results in successful outcomes (e.g., Anderson et al., 2005; Wallach et al., 2009).

Very few studies have investigated the effects of VR on high school students' public speaking performance and anxiety (Kahlon et al., 2019; Valls-Ratés et al., 2022; Valls-Ratés et al., 2023); and to our knowledge only one has explored adolescents' vocal and gestural patterns/behavior while being immersed in a VR scenario (see Valls-Ratés et al., 2021).

The aim of the present study was to evaluate adolescents' vocal and gestural patterns in virtual scenarios. Studies that have analyzed the effects of VR on prosody have been conducted in higher-education (e.g., Niebuhr & Michalsky, 2018; Notaro et al., 2021) and results show that VR can be an effective tool to develop more audience-oriented voices while performing a speech, pointing to a potential effect of VR-assisted public-speaking practice on not only the self-assessed anxiety of the participants but also their delivery style in terms of prosody and gesture use while practicing their speeches. To this end, in a mixed-model experimental design, high school students were asked to practice their speeches either in front of a VR audience (experimental condition) or alone in a classroom (control condition). The two conditions served as our between-subjects independent variable. Importantly, in order to have reference values in the three domains of interest, namely anxiety, prosody and gesture use, before rehearsing in one of the two conditions each speaker performed a short oral speech in front of a real audience of three people. The real vs. VR audience was a further within-subjects independent variable in our experiment.

We hypothesized that practicing speeches within VR settings would be conducive to (1) self-reports of higher levels of self-perceived anxiety in comparison to practicing alone without VR and (2) a more audience-oriented nonverbal communication style in terms of prosodies and gestures. In order to address the second part of the hypothesis, a comprehensive analysis of the 21 prosodic characteristics of the target speeches including pitch (i.e., f0), tempo and voice quality. Additionally, speakers' gesture rate during their public-speaking performance was determined.

3 METHOD

3.1 Participants, Experimental Design, Materials and Procedure

A total of 47 secondary school students from Barcelona participated in the study. The mean age of participants (67.18% female; 32.82% male) was 16.45 years (SD = 0.36). The experiment consisted of a mixed-model experimental design whose main factor was a between-subjects variable with the two conditions VR and Non-VR, see Figure 1. First, all the participants took part in a one-hour initial information session. Second, they performed two public speaking tasks, namely (a) a baseline two-minute public speaking task in front of a live audience and (b) a two-minute public speaking task, performed under one of two conditions, either in front of a VR-simulated audience (VR PRACTICE), or speaking alone (NON-VR PRACTICE). Following the information session, the experimenter randomly divided participants into two groups, the VR group (n = 27) and the Non-VR group (n = 20).

Figure 1: Experimental design.

3.1.1 Life Audience Speech (Baseline)

One week after the information session, participants were given written instructions on how to prepare what they planned to say, starting from the topic “Adolescents need to spend more time in nature” – and a list of five arguments that they could use and elaborate on in their speech. They were allotted two minutes for preparation and were then proceeded to the adjacent room where their audience was waiting. They were allowed a maximum of two minutes to deliver their speech.

3.1.2 VR and Non-VR Practice Sessions

The practice session took place one week after participants spoke to the live audience. The procedure for the practice was the same as for the speech to the live audience except that in this case the topic was “The house of my dreams" and they were offered a set of five questions instead of arguments to help them prepare the presentation. Again, after two minutes of preparation, they were accompanied to an adjacent classroom. At this point, however, the procedure followed diverged, depending on the group to which participants had been allocated, VR or Non-VR. Participants in the VR practice group were fitted with a Clip Sonic® VR headset, to which a smartphone was attached. A VR interface application installed on the smartphone called *BeyondVR* simulated a stage and gave the headset wearer the illusion that they were standing in front of an audience, see Figure 2. This virtual audience made small realistic movements while seated, and it conveyed an attentive attitude by making eye contact with the speaker and, more generally, signaling their interest in what the speaker was saying. These realistic features were intended to make the audience seem believably real and enhance the headset wearer’s sense of presence (Slater et al., 1999). VR group participants were able to monitor their speaking time by referring to a timer displayed in their field of vision by the headset. For Non-VR group participants, the procedure was the identical, except that they gave their speech alone in the classroom without any VR equipment. However, they had access to their speaking time on a computer screen placed close to them. The performance of all participants was video-recorded.

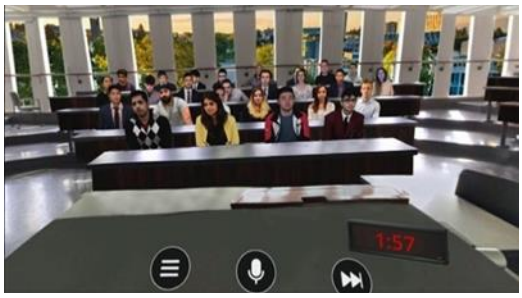

Figure 2: Screenshot of the VR scenario generated by BeyondVR that was seen by VR group participants.

3.1.3 Self-Assessed Anxiety

Just prior to performing the two public speaking tasks, each participant completed the Subjective Units of Distress Scale (SUDS) (Wolpe, 1969) form to indicate their level of anxiety. This instrument (a long-standing and validated questionnaire (see Thoen et al., 2023) yields a score from 0 (total relief) to 100 (the highest fear ever experienced). The participant was told "Please rate your level of distress from 0 to 100" and asked to read the descriptors for each 0-100 value in order to quantify their overall distress.

3.2 Data Analysis

Because each of the 47 participants delivered two speeches, one to the live audience and one in the practice session, a total of 94 recordings were obtained for analysis.

3.2.1 Prosodic Measures

Acoustic-prosodic analysis of the audio tracks of all 94 speeches was performed automatically by means of the ProsodyPro script by Xu (2013) and the supplementary analysis script by De Jong and Wempe (2009), both using the (gender-specific) default PRAAT settings (Boersma & Weenink, 2007). The analysis included a total of 21 different prosodic parameters, namely five f0 parameters, seven duration parameters and nine voice quality parameters. For a detailed explanation of the measured parameters see Liu & Xu, 2014.

3.2.2 Manual Gesture Rate

All manual communicative gestures present in the speakers’ speeches (to the live audience speech and in the practice speeches in the two conditions) were annotated. Following the M3D approach (see Rohrer et al., 2021 for more details on the procedure), we considered each manual stroke (the most effortful part of the gesture that usually constitutes its semantic unit; Kendon, 2004; McNeill, 1992) as corresponding to a manual gesture. Non-communicative gestures such as self-adaptors (e.g., scratching, touching hair; Ekman & Friesen, 1969) were disregarded. For every speech, the overall gesture rate was calculated as the total number of manual gestures produced relative to the phonation time in minutes (gestures/phonation time).

3.3 Statistical Analyses

Statistical analyses were performed using IBM SPSS Statistics version 19. A total of 23 of GLMMs were run for each of the following dependent variables: self-perceived anxiety (SUDS), a set of 21 values for all the prosodic parameters (five for f0, seven for duration and nine for voice quality), and the gesture rate. All the GLMM models included Condition (two levels: VR vs. Non-VR, between-subjects factor) and Time (two levels: Live Audience Speech; Practice Session) and their interactions as fixed factors. Subject was included as a random factor. Pairwise comparisons and post-hoc tests were carried out for the significant main effects and interactions.

4 RESULTS

4.1 Self-Assessed Anxiety

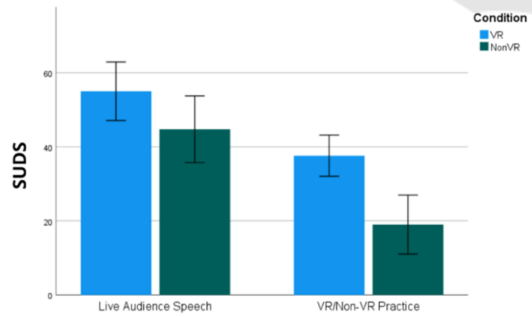

Figure 3: Mean SUDS values prior to the live audience speech and practice session for both VR and Non-VR conditions.

The GLMM analysis for SUDS showed a main effect of Condition (*F(1,88) = 13.513, p < .001*) that indicated that the participants of the VR group displayed significantly higher values than the Non-VR group, and not only for the practice speech but also when both groups were compared, respectively, to their live audience speeches (*β = 13.942, SE = 3.793, p < .001*). The analysis also showed a main effect of Time (*F(1,88) = 38.796, p < .001)*, meaning that the SUDS anxiety values obtained prior to the live audience speech (LAS) were significantly higher than those obtained prior to the practice session (PS) (β = 20.712, SE = 3.325, *p* < .001). Figure 3 shows mean SUDS scores separated by Condition (VR; Non-VR) and Time (Live Audience Speech = LAS; Practice Session = PS).

4.2 Prosodic Parameters

4.2.1 F0 Domain

Regarding the f0 domain, five GLMMs were applied to our target variables, namely minimum and maximum f0, f0 variability (in terms of the standard deviation), mean f0 and f0 range. Table 1 shows a summary of those GLMM analyses in terms of main effects (Time and Condition), as well as interactions between Time and Condition of the most relevant variables. Summarizing, a main effect of Time was obtained for f0 min, f0 variability and mean f0, showing that at Live Audience Speech values were higher for f0 min and mean f0, but significant lower for f0 variability. A main effect of Condition was obtained for f0 min, f0 max and mean f0, showing higher f0 values in the VR condition for the three variables. However, no significant interactions were obtained.

Table 1: Summary of the GLMM analyses for the f0 variables in terms of main effects and interactions (*ø* means no significant effect, whereas all non-*ø* table cells refer to significant differences between the paired conditions at p<0.05. Live Audience Speech = LAS; Practice Session = PS).

Variable	Main effect of Time	Main effect of Condition	Interaction Time*Condition
f0 min	*LAS > PS*	*VR > Non-VR*	*ø*
f0 max	*ø*	*VR > Non-VR*	*ø*
mean f0	*LAS > PS*	*VR > Non-VR*	*ø*

4.2.2 Tempo Domain

With regard to the tempo domain, seven GLMMs were applied to each of the target dependent variables, namely total number of syllables, total number of silent pauses, total time of the presentation, total speaking time, the speech rate, the net syllable rate and ASD. Table 2 shows a summary of these GLMM analyses in terms of main effects (Time and

Condition), as well as interactions between Time and Condition of the most relevant variables. Summarizing, no main effects of Time were obtained. A main effect of Condition was obtained for three variables: speech rate, net syllable rate and ASD, meaning that the participants in the VR group had significantly higher values for speech rate and net syllable rate values, and lower ASD values than the Non-VR group. However, no significant interactions were obtained.

Table 2: Summary of the GLMM analyses for the seven duration variables, in terms of main effects and interactions.

Variable	Main effect of Time	Main effect of Condition	Interaction Time*Condition
Number of silent pauses	ø	ø	ø
Speech rate	*LAS > PS*	*VR > Non-VR*	ø

4.2.3 Voice Quality Domain

Table 3: Summary of the GLMM analyses for the nine voice variables in terms of main effects and interactions.

Variable	Main effect of Time	Main effect of Condition	Interaction Time*Condition
h1*-h2*	ø	*VR > Non-VR*	*VR > Non-VR*
h1-A3	ø	*VR > Non-VR*	*VR > Non-VR*
CoG	*LAS < PS*	*VR > Non-VR*	ø
Formant dispersion 1-3	ø	*Non-VR > VR*	ø
Shimmer	ø	*VR > Non-VR*	*VR > Non-VR*
Jitter	ø	*VR > Non-VR*	*VR > Non-VR*
Hammarberg	*LAS < PS*	*VR > Non-VR*	*VR > Non-VR*

In the domain of voice quality measurements, nine GLMMs were applied to the nine target variables, namely h1*-h2*, h1-A3, CPP, Harmonicity, CoG, formant dispersion 1-3, shimmer, jitter, and Hammarberg index. Table 3 shows a summary of those GLMM analyses in terms of main effects (Time and Condition), as well as interactions between Time and Condition of the most relevant variables.

Summarizing, a main effect of Time was obtained for two variables, namely CoG and Hammarberg index, meaning that at baseline (Live Audience Speech) values were lower for CoG and higher for Hammarberg index. A main effect of Condition was

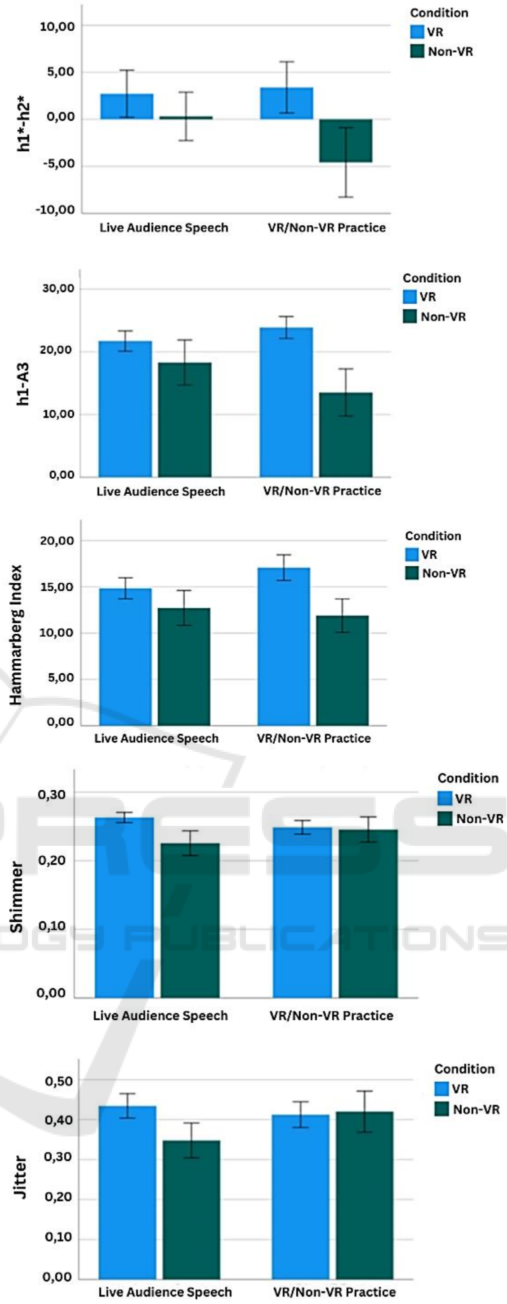

Figure 4: Mean voice quality values (namely h1*-h2*, h1-A3, Hammarberg index, jitter and shimmer) in the Live Audience Speech and the Practice Session for the variables that obtained a significant interaction Time * Condition, for both VR and Non-VR conditions.

obtained for seven variables, namely h1*-h2*, h1-A3, CoG, formant dispersion 1-3, shimmer, jitter and Hammarberg index, meaning that the participants in the VR group obtained higher values compared to the Non-VR group, in both the Live Audience Speech and the Practice Session, except for formant dispersion 1-3. Significant interactions were obtained for h1*-h2*, h1-A3, shimmer, jitter and Hammarberg index, showing higher values for the VR condition for all the variables. The graphs in Figure 4 show the mean voice quality values that obtained a significant interaction Time * Condition, for both VR and Non-VR conditions.

4.2.4 Manual Gesture Rate

A GLMM was applied for manual gesture rate. A main effect of Time was obtained ($F(1,84) = 40.601$, $p < .001$), showing that Live Audience Speech scores were higher across groups ($\beta = 16.410$, SE = 2.575, $p < .001$). However, no interaction was obtained between Time and Condition. The graph in Figure 5 shows the mean gesture rate values for both VR and Non-VR conditions.

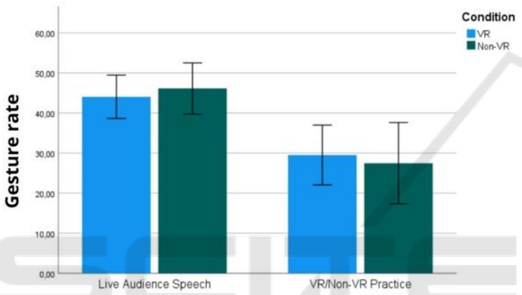

Figure 5: Mean gesture rate values for the Live Audience Speech and Practice Session for both VR and Non-VR conditions.

5 DISCUSSION AND CONCLUSIONS

The present study was designed to determine the effects of practicing a short oral presentation to an artificial VR-generated audience as compared to practicing alone in a classroom on self-perceived anxiety and a comprehensive set of prosodic features, together with gesture rate measures. Forty-seven high-school students participated in this mixed-model experiment. In order to obtain a baseline measure, all participants were asked to perform a speech in front of a live audience before performing the actual experimental task.

With respect to the first research question, it was found that the self-reported anxiety levels decreased significantly for both groups from the live audience speech to the practice session. No significant interactions were found, meaning that the two conditions were not different with respect to the baseline. The lack of effect on self-reported anxiety measures could be explained by two reasons. First, SUDS self-reports were the only measure of participants' distress arousal prior to the performance of public speaking. Adding other questionnaires and combining them with physiological measures would have allowed us to obtain a more fine-grained picture of the distress and anxiety levels as the speakers faced and talked to live and virtual audiences. Also, it is conceivable that higher stress levels would be reported *after* participants had put on the headset and were facing the virtual audience than those reported before the headset was put on.

Turning to the prosodic analysis of the participants' speeches, and to the tempo domain at first, no significant interactions were obtained for any of the parameters (total number of syllables, total number of silent pauses, total time of the presentation, total speaking time, speech rate, net syllable rate and ASD), meaning that both VR and Non-VR groups showed a similar (change in) behavior when their practice speeches are compared to their live audience speeches. In the f0 domain, a main effect of time was obtained for f0 min, f0 variability, mean f0 and range, and a main effect of condition for f0 min, f0 max, and mean f0. The main effect of Condition revealed that f0 min, f0 max and mean f0 were higher for the VR condition than for the Non-VR condition. These results are in line with those reported by Niebuhr and Michalsky (2018), Notaro et al. (2021) and Remacle et al. (2020), who found higher f0 levels and higher-level melodic variation when participants were immersed in artificial VR environments. In Niebuhr and Michalsky, however, there is no comparison to a speech performed in front of a live audience, unlike in the other two studies (Notaro et al., 2021 and Remacle et al., 2020) as well as the present study. However, we see that all f0 parameters maintain the same high levels in the live audience speech as in the practice session. Thus, no significant interaction was found here between time and condition, meaning that there was no significant difference between the f0 characteristics of speech in the VR and the Non-VR conditions relative to the f0 features of the baseline speech condition in front of an audience.

Importantly, the main difference between this study and previous ones is the fact that our study did not include any form of feedback during or after the VR practice. Providing feedback on nonverbal aspects of a speaker's performance seems to be fundamental to achieving improvement, as VR per se does not include this feature in an automatic way (e.g., Niebuhr & Michalsky, 2018). Importantly, our study did not find a significant difference between the VR and non-VR conditions in the self-assessment measure of anxiety, as well as in the duration and f0 measures of practice speeches. However, even though this result distinguishes our study from some previous findings, it needs to be related to the fact that we have taken into account baseline measures. We believe that a between-subjects factor involving VR-assisted and non-VR-assisted conditions must be checked against such within-subjects baseline condition where participants perform a speech before a live audience in order to assess potential individual differences.

Crucially, in the voice quality domain (h1*-h2*, h1-A3, CPP, Harmonicity, CoG, formant dispersion 1-3, shimmer, jitter, and Hammarberg index), significant interactions were obtained between time and condition. Specifically, VR-assisted speakers tended to use a louder and more powerful voice than Non-VR-assisted speakers relative to when they were addressing a live audience speech. These results are consistent with Niebuhr & Michalsky (2018) and Remacle et al. (2020). Practicing with VR also was reflected in a higher Hammarberg index and h1-A3, which suggests a more effortful and aroused voice quality (Niebuhr & Taghva, 2022; Tamarit et al., 2008). VR speeches showed a significant decrease for shimmer, that is, less shaky, nervous, stressful voice, whereas the opposite was found for the Non-VR speeches. By contrast, Non-VR participants significantly increased in jitter during their practice session speeches, showing a less harmonic, tenser or creaky voice.

Regarding the effects of manual gesture rate, when comparing the amount of gesture produced in front of the live audience compared to the practice session, there was a significant reduction in the gesture rate for both groups. That is, VR and Non-VR participants both produced gestures less often in the practice sessions. This decrease was greater in the Non-VR condition, albeit not significantly so. These results are in line with Notaro et al. (2021). Following their reasoning, a possible explanation for this might be that participants who are immersed in VR and wearing a headset cannot see their own hands. For Non-VR group participants, the reason was probably different; here, the decrease may be due to an absence of motivation and engagement because they are alone in a classroom, giving a speech to no one. Taking these results into account, we cannot confirm the hypothesis that predicted an increase in gesture rate in the VR condition, as the tendency was very similar across the two groups.

In summary, participants that practiced their short speeches within an unsupervised VR environment in front of a virtual audience had in effect a more realistic experience (in line with Selck et al., 2022). As a result, regarding prosodic parameters, we see an increase in vocal effort and loudness, with voices that are stronger (hence also less shaky and stressed) and aroused, which reflects a more audience-oriented manner of speaking. The presence of the virtual audience made participants more engaged and encouraged them to use their voices similarly to how they would have done in front of a live audience.

Some limitations of the study must be noted. First, the study is based on a relatively small sample size (for having a between-subjects factor involved). So, results cannot be generalized to other age groups or clinical populations. Second, assessing gesture rate might not be enough to differentiate between the gestural behavior of participants in both conditions. Adding a more complete assessment of overall multimodal behavior including body movement, facial expressions, eye contact, types of hand gesture, and so on could expand our knowledge about the participants' body engagement during VR experiences. Finally, the addition of feedback about speaker performance during or after public speaking performances could have also favored the public speaking VR experience of these young students, as feedback has been shown to be valuable for learning and skills improvement (Van Ginkel et al., 2019; King et al., 2000).

There is abundant room for further progress in determining the effects of VR immersion in students' overall performances not just in one practice session, but in several, to explore the differences between the first VR immersion and the subsequent ones. Also, not only comparting VR environments to being alone in a classroom, but also comparing it to other ways of rehearsing a speech. Longer VR sessions as well as longer time to prepare the speeches could be also further explored.

All in all, the present study highlights the value of using VR for public speaking practice in secondary school settings. If the current trend is for educational policies to promote the learning of public speaking skills, then opportunities should be provided for students to rehearse their presentations and speeches using virtual environments. In our view, combining VR immersion with other sorts of training in the classroom to develop related skills such as quality conversation, active listening and critical thinking can be key to broadening students' competence in both their daily and future professional lives, in a more engaging and fun way.

Future research should focus on analyzing the long-term effects of virtual simulations on real environments. Longitudinal studies could assess students' perceptions of enjoyment and usefulness with VR to determine if their high regard for it persists beyond novelty. Additionally, extending training sessions, duration, and incorporating feedback strategies are essential future aims in both research and practice.

Communicating effectively is an ability that needs to be developed and mastered through the years. The earlier we start teaching these skills, the easier it will be to apply them properly in any communicative situation. By nurturing these skills early on, educators and parents empower young learners to communicate effectively, collaborate productively, and navigate the world with confidence and competence.

REFERENCES

Anderson, P. L., Zimand, E., Hodges, L. F., & Rothbaum, B. O. (2005). Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and Anxiety, 22*(3), 156-158.

Baños, R. M., Guillen, V., Quero, S., García-Palacios, A., Alcañiz, M., & Botella, C. (2011). A Virtual Reality System for the Treatment of Stress-Related Disorders: A Preliminary Analysis of Efficacy Compared to a Standard Cognitive Behavioral Program. *International Journal of Human-Computer Studies, 69*(9), 602-613.

Bianchi-Berthouze, N. (2013). Understanding the Role of Body Movement in Player Engagement, *Human-Computer Interaction, 28*(1), 40-75.

Boersma, P., & Weenink, D. (2007). PRAAT: Doing Phonetics by Computer (Version 5.2.34). [Computer software].

Bourdin, P., Martini, M., & Sanchez-Vives, M. V. (2019). Altered visual feedback from an embodied avatar unconsciously influences movement amplitude and muscle activity. *Scientific Reports, 9*(1), [19747]. doi: 10.1038/s41598-019-56034-5

Daniels, M. M., Palaoag, T., & Daniels, M. (2020). Efficacy of virtual reality in reducing fear of public speaking: A systematic review. *In International Conference on Information Technology and Digital Applications, 803*(1). doi: 10.1088/1757-899X/803/1/012003

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior research methods, 41*(2), 385-390.

Ekman, P., & Friesen, W.V. (1969). The Repertoire or Nonverbal Behavior: Categories, Origins, Usage and Coding. *Semiotica, 1*, 49-98. doi: 10.1515/semi.1969.1.1.49

Ford, W. S. Z., & Wolvin, A. D. (1993). The differential impact of a basic communication course on perceived communication competencies in class, work, and social contexts. *Communication Education, 42*, 215-233.

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers, *Cognition, 177*, 263-277. doi:10.1016/j.cognition.2018.04.007.

Huang, B. H., & Jun, S. A. (2011). The Effect of Age on the Acquisition of Second Language Prosody. *Language and Speech, 54*(3), 387-414. doi: 10.1177/0023830911402599

Hui Lim, M., Aryadoust, V., & Esposito, G. (2022). A meta-analysis of the effect of virtual reality on reducing public speaking anxiety. *Current Psychology*. Advance online publication. doi: 10.1007/s12144-021-02684-6

Iberri-Shea, G. (2009). Using Public Speaking Tasks in English Language Teaching. *English Teaching Forum, 47*(2), 18-23.

Kaufmann, R., & Tatum, N.T. (2017). Do we know what we think we know? On the importance of replication in instructional communication research. *Communication Education, 66*, 479-481.

Kahlon, S., Lindner, P., & Nordgreen, T. (2019). Virtual reality exposure therapy for adolescents with fear of public speaking: A non-randomized feasibility and pilot study. *Child and Adolescent Psychiatry and Mental Health, 13*(47), 1-10.

Kendon, A. (2004) Gesture: Visible Action as Utterance. Cambridge University Press, Cambridge.

King, P., Young, M., & Behnke, R. (2000). Public speaking performance improvement as a function of information processing in immediate and delayed feedback interventions. *Communication Education, 49*(4), 365-374. doi: 10.1080/03634520009379224

Liao, H. (2014). Examining the Role of Collaborative Learning in a Public Speaking Course. *College Teaching, 62(2)*, 47-54, doi: 10.1080/87567555.2013.855891

Liu, X. & Xu, Y. (2014). Body size projection by voice quality in emotional speech-Evidence from Mandarin Chinese. *Proceedings of the International Conference on Speech Prosody*, 974-977.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* University of Chicago Press.

Morreale, S. P., Osborn, M. M., & Pearson, J. C. (2000). Why communication is important: A rationale for the centrality of the study of communication. *Journal of the Association for Communication Administration, 29*(1), 1-25.

Munby, I. (2011). The oral presentation: An EFL teachers' toolkit. *Studies in Culture, 99*, 143-168.

Nguyen, H. (2015). Student perceptions of the use of Pecha Kucha presentations for EFL reading classes. *Language in Education in Asia, 6*(2), 135-149.

Niebuhr, O. & Taghva, N. (2022, in review). Posture and Gestures can affect the prosodic speaker impact in a remote presentation. In M. Oliveira (Ed.), *Prosodic Interfaces*. Berlin/New York: de Gruyter.

Niebuhr, O. & Michalsky, J. (2018). Virtual reality simulations as a new tool for practicing presentations and refining public-speaking skills. In Proceedings of the 9th International Conference on Speech Prosody 2018, 309-313. *International Speech Communication Association* (ISCA).

Notaro, A., Capraro, F., Pesavento, M., Milani, S. & Busà, M.G. (2021). Effectiveness of VR immersive applications for public speaking enhancement. *In Proc. IS&T Int'l. Symp. on Electronic Imaging: Image Quality and System Performance XVIII, 2021, 294-1 - 294-7*. doi: 10.2352/ISSN.2470-1173.2021.9.IQSP-294

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, *147*, [103778]. doi: 10.1016/j.compedu.2019.103778

Remacle, A., Bouchard, S., Etienne, A., Rivard, M., & Morsomme, D. (2021). A virtual classroom can elicit teachers' speech characteristics: evidence from acoustic measurements during in vivo and in virtuo lessons, compared to a free speech control situation. *Virtual Reality, 25*(4), 935-944.

Rohrer, P. L., Vilà-Giménez, I., Florit-Pons, J., Gurrado, G., Gibert, N. E., Ren, A., Shattuck-Hufnagel, S., & Prieto, P. (2021, February 24). The MultiModal MultiDimensional (M3D) labeling system. doi: 10.17605/OSF.IO/ANKDX

Schneider, J., Borner, D., van Rosmalen, P., & Specht, M. (2017). Presentation Trainer: what experts and computers can tell about your nonverbal communication. *Journal of Computer Assisted Learning, 33*(2), 164-177. doi: 10.1111/jcal.12175

Selck, K., Albert, T., & Niebuhr, O. (2022). "And miles to go before…"–Are speech intensity levels adjusted to VR communication distances?. In *Book of Abstracts of the 13th Nordic Prosody Conference, Sønderborg, Denmark,* 44-46.

Slater, M., Pertaub, D. P., & Steed, A. (1999). Public speaking in virtual reality: Facing an audience of avatars. *IEEE Computer Graphics and Applications, 19(*2), 6-9.

Tamarit, L., Goudbeek, M., & Scherer, K.R. (2008). Spectral slope measurements in emotionally expressive speech. *In proceedings of SPKD-2008, paper 007.*

Thoen, A., Steyaert, J., Alaerts, K., Evers, K., & Van Damme, T. (2023). A Systematic Review of Self-Reported Stress Questionnaires in People on the Autism Spectrum. Review *Journal of Autism and Developmental Disorders, 10*(2), 295-318. doi: 10.1007/s40489-021-00293-4

Valls-Ratés, Ï., Niebuhr, O., & Prieto, P. (2021). Effects of public speaking virtual reality trainings on prosodic and gestural features. *Proc. 1st International Conference on Tone and Intonation (TAI). University of Southern Denmark*. doi: 10.21437/TAI.2021-44

Valls-Ratés, I., Niebuhr, O., & Prieto, P. (2022). Unguided virtual-reality training can enhance the oral presentation skills of high-school students. *Frontiers in Communication, 7,* [910952]. doi: 10.3389/fcomm.2022910952

Valls-Ratés, I., Niebuhr, O., & Prieto, P. (2023). Encouraging participant embodiment during VR-assisted public speaking training improves persuasiveness and charisma and reduces anxiety in secondary school students. *Frontiers in Virtual Reality, 4,* [1074062]. doi: 10.3389/fvir.2023.1074062

Van Ginkel, S., Ruiz, D., Mononen, A., Karaman, A. C., de Keijzer, A. & Sitthiworachart, J. (2020). The impact of computer-mediated immediate feedback on developing oral presentation skills: An exploratory study in virtual reality. *Journal of Computer Assisted Learning, 36.* doi: 10.1111/jcal.12424

Wolpe, J. (1969). *The Practice of Behavior Therapy*, Pergamon General Psychology Series, 1. 4th Edition. Elmsford, NY, US: Pergamon Press.

Xu, Y. (2013). ProsodyPro - A Tool for Large-scale Systematic Prosody Analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody* (TRASP 2013), Aix-en-Provence, France. 7-10.