



OPEN ACCESS

EDITED BY

Sao Luis Castro,
University of Porto, Portugal

REVIEWED BY

Weimin Toh,
Nanyang Technological
University, Singapore
Likun Zhan,
Beijing Language and Culture
University, China

*CORRESPONDENCE

Ío Valls-Ratés
io.valls@upf.edu

†These authors have contributed
equally to this work

SPECIALTY SECTION

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

RECEIVED 01 April 2022

ACCEPTED 23 August 2022

PUBLISHED 30 September 2022

CITATION

Valls-Ratés Í, Niebuhr O and Prieto P
(2022) Unguided virtual-reality training
can enhance the oral presentation
skills of high-school students.
Front. Commun. 7:910952.
doi: 10.3389/fcomm.2022.910952

COPYRIGHT

© 2022 Valls-Ratés, Niebuhr and
Prieto. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Unguided virtual-reality training can enhance the oral presentation skills of high-school students

Ío Valls-Ratés ^{1*}†, Oliver Niebuhr ^{2†} and Pilar Prieto ^{1,3†}

¹Department of Translation and Language Sciences, Universitat Pompeu Fabra, Barcelona, Spain,

²Centre for Industrial Electronics, University of Southern Denmark, Sønderborg, Denmark,

³Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Public speaking is fundamental in our daily life, and it happens to be challenging for many people. Like all aspects of language, these skills should be encouraged early on in educational settings. However, the high number of students per class and the extensive curriculum both limit the possibilities of the training and, moreover, entail that students give short in-class presentations under great time pressure. Virtual Reality (VR) environments can help speakers and teachers meet these challenges and foster oral skills. This experimental study employs a between-subjects pre- and post-training design with four Catalan high-school student groups, a VR group ($N = 30$) and a Non-VR group ($N = 20$). Both groups gave a 2-min speech in front of a live audience before (pre-training) and after (post-training) 3 training sessions (one session per week) in which they practiced public speaking either in front of a VR audience or alone in a classroom (Non-VR). Students assessed their anxiety measures right before performing every speech and filled out a satisfaction questionnaire at the end. Pre- and post-training speeches were assessed by 15 raters, who analyzed the persuasiveness of the message and the charisma of the presenter. Speeches were also analyzed for prosodic features and gesture rate. First, results showed that self-assessed anxiety got significantly reduced at post-training for both conditions. Second, acoustic analyses of both groups' speeches show that the VR group has, unlike the Non-VR group, developed a more clear and resonant voice quality in the post-training speeches, in terms of higher cepstral-peak prominence (CPP) (although no significant differences in f0-related parameters as a function of training were obtained), as well as significantly less erosion effects than the Non-VR group. However, these differences across groups did not trigger a direct improvement on the participants' gesture rate, persuasiveness and charisma at post-training. Furthermore, students perceived the training in the VR environment to be more useful and beneficial for their future oral presentations than the Non-VR group. All in all, short unguided VR training sessions can help students feel less anxious, promote a more clear and resonant voice style, and can prevent them from experiencing an erosion effect while practicing speeches in front of a real audience.

KEYWORDS

public speaking skills, gesture, virtual reality, anxiety, charisma, prosody

Introduction

Boosting public speaking abilities in secondary school settings contributes not only to strengthening students' effectiveness with academic work (cf. the anecdote in Fox Cabane, 2013, pp. 139–141), but also their social skills, thus affording them more satisfactory interpersonal relationships (e.g., Morreale et al., 2000; Bailey, 2018) and preventing them from abandoning their studies prematurely (e.g., Boettcher et al., 2013; Niebuhr, 2021). In order to achieve these goals, it would be desirable that high schools acknowledge the importance of oral abilities for enhancing students' self-confidence and that they take action by involving students more often in oracy settings that encourage them to actively take part in their community (Bailey, 2018). However, time restrictions and the pandemic situation make it difficult for teachers to organize oral practices in front of the classroom. The present paper assesses the use of virtual reality technology (henceforth VR) as an alternative and complementary educational method for practicing oral presentations. Given the fact that VR can easily simulate traditional training scenarios in a virtual environment, the present investigation will determine the effects of a short 3-session VR training with high school students on reducing their public speaking anxiety and enhancing the quality of their oral presentations after training.

The importance of public speaking practice in educational settings

As any other skill, public speaking needs practice. One of the widely used instruction techniques in the educational system is the delivery of oral presentations by students, as they are frequently asked to present their projects or research papers in front of their peers. Yet one of the problems students face with this type of task is the fear of public speaking. PSA (or Public Speaking Anxiety, also called glossophobia) is related to different physiological changes like elevated heart and breathing rates, over-rapid reactions, trembling of muscles and shoulder and neck area stiffness (Tse, 2012). High levels of PSA can result in poor speech preparation (Daly et al., 1995) and impede decision-making of effective speech introduction strategies (Beatty and Clair, 1990; Beatty, 1998). Also, highly anxious individuals may be perceived by the audience as more nervous, they make less eye contact and pause more often than less anxious individuals (Daly et al., 1995; Choi et al., 2015); and most obviously, the quality of their speech performance is negatively affected (Beatty and Behnke, 1991; Menzel and Carrell, 1994; Brown and Morrissey, 2004). The negative thinking of those speakers exhibiting larger levels of PSA can reduce their speaking competence (Daly et al.,

1995; Rubin et al., 1997, and make them procrastinate in speech preparation (Behnke and Sawyer, 1999).

In practice, PSA and speech delivery problems can be effectively addressed by offering students more opportunities to rehearse their oral presentations. Goberman et al. (2011) showed that the earlier speakers started rehearsing their presentations on their own (i.e., unguided), the more fluent their speeches were after practicing, but with narrower pitch variation ranges compared with the students who started practicing later. A similar “prosodic erosion” effect (a successive lowering and narrowing of their speech melody across the repeated rehearsals of their presentation) is reported by Niebuhr and Michalsky (2018) (see also Niebuhr and Tegtmeier, 2019). Importantly, research shows that oral skills practice optimally needs to be performed orally in front of an audience. Smith and Frymier (2006) found that, compared with students rehearsing alone, rehearsing in front of an audience gave students higher scores on their final classroom-speech assessment, thus lending support to the claim that audience-based speech practice can help increase public speaking performance. Menzel and Carrell (1994) showed that practicing oral presentations before a classroom audience is the single greatest predictor of student speaking success and key for reducing PSA.

However, organizing such a setup can be difficult for teachers, given the high number of students per class and the extensive curriculum that needs to be covered in courses. The situation has been aggravated with the pandemic situation, where face-to-face interaction was limited to a great extent. Moreover, a high percentage of students dedicate most of their time to writing their speech rather than to rehearsing it orally, spending an average of <5 min on oral rehearsing (see Pearson et al., 2006). Given this situation, in the following section we assess the previous literature on the value of using VR as a complementary educational tool for providing an appealing setup for practicing audience-based oral presentations and thus boosting public speaking skills.

A complementary solution: Empirical evidence on the effects of VR for boosting public speaking skills

As a way to enhance the oral practice of presentations and, also, to reduce anxiety when delivering speeches in front of an audience, VR simulations can be of great help. Virtual simulations can be broadly defined as 3D interactive environments that are computer-generated and are viewed by a single user through a headset that excludes all other visual input. While many of these VR platforms have been traditionally used for entertainment purposes, a large number of schools, hospitals, and research institutions (Peeters, 2019) are currently using this technology to provide active learning environments

(Legault et al., 2019). Since VR experiences evoke realistic responses in people, they can be fundamentally conceived as “reality simulators.” Participants in VR settings are placed in an artificial scenario that depicts potentially real events, with the likelihood that they will act and respond realistically. VR gives rise to the subjective illusion that is referred to in the literature as *presence*—the illusion of “being there” in the environment depicted by the VR displays—in spite of the fact that the user is simultaneously fully aware that the environment is artificial (Armel and Ramachandran, 2003). VR is different from other forms of human–computer interface “since the human participates in the virtual world rather than uses it” (Slater and Sanchez-Vives, 2003, p. 3). Mikropoulos and Natsis’s (2011) empirical study dealing with the application of virtual reality in learning environments suggests that “presence is considered to be a key feature” with a majority of the practitioners, whose work they examined, reporting that “their sample had the feeling of ‘being there’ and that this might contribute to positive results” (p. 774). Accordingly, “being there” leads to the participants’ increase in “intrinsic motivation and engagement” (Dalgarno and Lee, 2010). Ruscella (2019) and LeFebvre et al. (2020) suggest that an immersive setting reduces fear and creates a no-risk situation that is ideal for learners to practice their speeches. As LeFebvre et al. (2020, p. 10) points out, “VR creates a more effective treatment environment for enacting changes to reduce PSA.” Even though information about public speaking might not be provided to the user, spending time practicing in front of the virtual audience may improve social skills that can be transferred to the real world (Xu et al., 2011; Lane et al., 2013; Rogers et al., 2017; Howard and Gutworth, 2020).

Effects of VR to treat public speaking anxiety

In the context of public speaking training, some studies have tested the use of VR technology to reduce anxiety in university students. In a systematic review, Daniels et al. (2020) identified 14 studies conducted from 2009 to 2019 that used VR as a tool to diminish public speaking anxiety (PSA). From these 14 studies, 7 belonged to clinical settings (Wallach et al., 2009, 2011; Lister et al., 2010; Lister, 2016; Lindner et al., 2018; Yuen et al., 2019; Zacarin et al., 2019). Three of the 7 clinical studies (Wallach et al., 2009, 2011; Lister et al., 2010) compared PSA levels before and after VR immersion and found a significant PSA reduction. Wallach et al. (2009) compared, with 88 participants, Cognitive Behavioral Therapy (CBT) to VR immersion in a total of 7 sessions, and they found that both treatments were effective in reducing speakers’ anxiety (see also (Safir et al., 2012)). In a later study, Wallach et al. (2011) applied the same design, with 20 female participants, this time comparing Cognitive Therapy (CT) to VR. They yielded the same results regarding both treatments. Lister et al. (2010), in a study with 20 participants, found that VR 3D videos were capable of eliciting a fear response in participants

and was effective in reducing negative self-beliefs about public speaking abilities. In the study by Lindner et al. (2018), with 50 participants, they compared therapist-led exposure followed by 4 VR internet intervention sessions to a self-led waiting list (WL) condition. They concluded that those internet interventions were as effective as the traditional therapist-led interventions in reducing speakers’ PSA. Moreover, VR intervention sessions showed that this cost-effective technology can lead to solid and promising automated self-help applications. In another study by Lindner et al. (2020) with 25 participants, they showed that only one session of VR exposure therapy constituted an effective treatment of PSA. Lister (2016), in a study with 98 participants that compared a VR condition to a control condition, concluded that six sessions were capable of increasing confidence of speakers and obtained positive self-statements. Two clinical studies included in the systematic review did not include control conditions, namely Yuen et al. (2019) and Zacarin et al. (2019). Yuen et al. (2019) in two pilot studies with 11 and 15 participants each, showed that 6 weekly sessions were enough to significantly reduce PSA in a 3-month follow-up test. In the study by Zacarin et al. (2019), with 6 female participants, they designed 6 individual sessions and 1- and 3-month follow-up sessions, all including feedback by the therapist. Results showed that feedback allowed them to improve their speech and that this contributed to reducing their anxiety. Also, an increase in speaking quality was found in terms of a reduction of silent pauses and of word repetitions.

The other 7 studies included in the systematic review were performed in university educational settings. Two of them compared PSA from pre to post treatment and found a significant reduction (Heuett and Heuett, 2011; Nazligul et al., 2017), whereas the other five had different research designs. Heuett and Heuett (2011) carried out a study with 80 university students. The pre-training sample gave an impromptu speech and filled out questionnaires related to PSA and Willingness to Communicate (WTC)—and was then randomly assigned to one of three groups. One group practiced public speaking to a VR-generated virtual audience, another group was trained to visualize an audience as they spoke, and the third group, i.e., the control group, received no training at all. Both treatments lasted between 10 and 20 min, after which all three groups carried out a post-test which was identical to the pretest, and all participants completed the same questionnaires again. A comparison of pre-training and post-test data from the participants in the VR group showed a significant reduction in trait and state communicative apprehension (CA), and an increase in their self-perceived communication competence (SPCC) and WTC scores. The visualization treatment also yielded significant improvements in trait and state CA and SPCC, but not in WTC. The control group reported no significant change for any of the variables studied. The other study, by Nazligul et al. (2017), was conducted with 6 software engineers university students (21 years old). Every participant attended a 1-h individual therapy session

where they were told about anxiety and its possible causes and components, and they rated their self-perceived anxiety level while imagining giving a speech. After that, they performed a brief speech on a controversial topic and rated their self-assessed anxiety with the SUDS at 4 different points during exposure. Participants reported that, while being exposed to VR, they felt the highest level of anxiety, but also lower levels of anxiety after the intervention ended. There was no control group.

Two other educational studies that had no control group were [Stupar-Rutenfrans et al. \(2017\)](#) and [Takac et al. \(2019\)](#). The former was conducted with 19 university students and demonstrated in a within-subject task design that rapidly successive VR scenarios could elicit self-reported distress, and significant physiological arousal was also observed in heart rate data. Distress was easier to trigger than habituation, with three successive speeches (within a 60-min session) required to sustain distress reduction. [Stupar-Rutenfrans et al. \(2017\)](#) carried out a further study in which 35 university students performed three different speeches, one per week, using VR technology at home. In the first session the VR screen showed no audience, in the second the VR screen showed a small audience and in the third, a large audience. Participants had to fill out three questionnaires to assess their levels of anxiety and emotion regulation during treatment: namely the Emotion Regulation Questionnaire ([Gross and John, 2003](#)), the Public Report of Communication Apprehension ([McCroskey, 1982](#)), and the STAI Inventory. The study concluded that initially more anxious participants significantly improved in self-assessed anxiety scores after having performed in all three VR conditions. Their anxiety increased between the first and second session but diminished before and after the third session. The authors recommended that future research in that line should include a control group and also pre- and post-training tasks that would include speaking to a live audience in order to compare the reduction of anxiety in virtual and non-virtual public speaking contexts.

[Aymerich-Franch and Bailenson \(2014\)](#), [North et al. \(2015\)](#), and [Wilsdon and Fullwood \(2017\)](#) conducted educational studies that included both a VR and a control condition. The former study had a total of 14 participants and compared VR (7 participants) to a no-treatment group (7 participants) in a total of 5 sessions. They found a significant reduction in fear measures in the treatment group, but no relative comparison between groups was made. [Aymerich-Franch and Bailenson \(2014\)](#), with a sample of 41 participants, conducted a study with a VR group that performed visualization with a doppelganger (virtual humans that highly resemble the real self but behave independently) and a control condition that performed visualization with imagination. For VR participants, the first part of the session consisted of seeing their doppelganger performing a successful speech through VR while listening to a relaxing voice. The control group had to imagine giving a successful speech while listening to the relaxing

voice. After that, participants of both groups performed a speech on a topic of their choice before an audience of two people. They concluded that there were no differences in self-perceived anxiety across groups. However, they found an interaction between condition and gender for state anxiety and self-perceived communicative competence. The doppelganger technique worked better for males, and as the authors point out this was probably because men were already more familiarized to be in virtual environments and felt more comfortable during the VR experience, whereas the visualization technique proved more effective for females. To our knowledge, only one study has reported null effects of VR training on anxiety. [Wilsdon and Fullwood \(2017\)](#) conducted a one-session study with 40 university students consisting of 3 VR conditions (high, medium, and low immersion environments) and a control condition. The VR conditions performed a 5-min speech about their first week at university before a VR audience, while the control condition performed the same speech to the researcher. Participants filled in anxiety self-assessment questionnaires before and after the speech task. Results showed no improvement in PSA reduction, and increased VR immersion did not significantly reduce their anxiety either.

Besides the studies included in the systematic review, there are other studies that also show positive results in anxiety reduction: [Harris et al. \(2002\)](#) in a study that involved 14 university students with a VR group and a WL group, found that four 15-min sessions of VR were effective for reducing PSA. The pre-training consisted of different short public speaking tasks and different self-report instruments. The VR group then underwent four training sessions with different tasks while the WL group was given the same VR training once the experimental data had been gathered. Post-testing consisted of the same respective tasks. Although there were significant reductions in anxiety at post-test on some measures in the VR group (self-assessed questionnaires and heart rate), only one comparison between the VR and the WL group proved to be significant—i.e., the one that compared levels of speaker self-confidence. VR participants showed greater improvement overall on both self-assessment and physiological measures. [Rodero and Larrea \(2022\)](#) conducted a study with 100 university students, and they were divided into a VR experimental group and a control group. They performed a pre-training and a post-training task which consisted of giving a 3-min speech in front of a live audience. Trainings consisted of 5 trial sessions with a VR environment for the experimental group, whereas for the control group the 5 training sessions were led by an instructor. During the training sessions in both conditions, the authors included distractors (someone coughing in the audience or someone in the audience asking a question). The study measured self-assessed anxiety and electrodermal activity. Results show that VR participants significantly reduced their anxiety levels (in both measures) and that distractors (someone coughing placed at second 40 and someone's question at second 60, in pre- and

post-test speeches) proved effective at reducing their anxiety at post-test. Therefore, they conclude that training with distractors is effective and reproduces a more real public speaking situation. Participants said that training with VR helped them concentrate, made them more confident and made them have less tension.

To our knowledge, only one study (Kahlon et al., 2019) has previously examined VR effects on PSA reduction in a secondary school setting. They studied the PSA of 27 adolescents (aged 13–16) after only a single 90-min VR session, in which they performed different speaking or public speaking exercises. Subsequently, they received brief psychoeducation, active maintenance and filled in different anxiety self-assessments. A therapist accompanied them throughout the session. The authors concluded that one session was enough to reduce PSA of adolescents after 1- and 3-month follow ups, although the causes for this PSA reduction are not clear as there were neither control nor comparison groups.

Effects of VR on students' motivation

All in all, there is evidence that VR serves as a tool to trigger anxiety during training sessions and eventually reducing anxiety after training. However, in the context of educational practice, are VR public speaking trainings capable of stimulating a higher commitment to learning, in particular with respect to high-school students as the target group?

Several studies have shown that students are highly motivated using VR technology for practicing public speaking. The study by Frisby et al. (2020) concludes that employing VR for speech rehearsals not only helps diminish PSA. Rather, students consider it an innovative way of oral rehearsing that makes them more willing to accomplish a good performance. Vallade et al. (2020) and Kryston et al. (2021) also report on the excitement of students to participate in VR experiments as a different and motivating way to entice them to rehearse their speeches. Specifically, Kryston et al. showed how participants in the VR settings reported that it was more demanding than other modes of practice, which is consistent with the ability of digital audiences to elicit mental stress in speakers. In their qualitative study, Gruber and Kaplan-Rakowski (2020) examined the efficacy of VR based on the perception of 12 university students performing 8 different speeches. They analyzed the participants' sense of presence, the plausibility of the illusion and the perceived usefulness of VR for practicing public speaking. Although the sample was small, participants acknowledged the potential of VR for practicing oral speeches, compared to traditional practices, they saw cognitive benefits of the VR experience and they would find it useful as a tool to practice oral presentations to be presented in front of university audiences. They also emphasized how practicing with VR made them more capable of speaking in front of live audiences. Findings by Daniels (2021) showed that the usability ratings of virtual reality as a training tool for public speaking training

can vary depending on the technological background of users. They concluded that “the use of virtual reality as a training tool for public speaking training is highly recommended. This is supported by the unanimously positive responses of participants in the System Usability Scale (SUS) that measures their interest in using the VR tool for oral presentations” (Daniels, 2021, p. 6).

Effects of VR as conducive of a more listener-oriented prosodic style

Given that VR provides a credible set of scenarios that allow for an immersive learning situation, when used for public speaking tasks, VR environments have been reported to be conducive to a more listener-oriented speaking style from the point of view of the prosodic characteristics. To our knowledge, five studies have assessed the impact of using VR on the speech characteristics of the speakers while using this technique during a public speaking task as compared to other conditions. Three of them (Niebuhr and Michalsky, 2018; Remacle et al., 2021; Valls-Ratés et al., 2021) put the focus on prosody (which refers to all aspects of a speaker's voice and tone-of-voice). Niebuhr and Michalsky (2018) showed in a study with 24 participants comparing VR and Non-VR groups, that those students rehearsing public speeches within a VR environment performed their speech in a more listener-oriented, conversation-like speaking style than participants in the control group, who practiced their speech alone in a classroom. They concluded that the speeches of participants who were trained in the VR condition were more charismatic and more audience-oriented (characterized by a higher F0 level, a larger F0 range, and a slower speaking rate), showing reduced signs of “prosodic erosion” due to repeated rehearsing, compared to those participants who had practiced their speeches alone in a classroom (see also Niebuhr and Tegtmeier, 2019). Moreover, compared to the control Non-VR group, the speakers were unexpectedly motivated to speak longer, and the speech of the VR group was characterized by higher fundamental-frequency (i.e., f0) levels, a wider f0 range, a slower speaking rate, fewer pauses and a higher intensity level. A recent study by Remacle et al. (2021) conducted with 30 female elementary school teachers also proved to be effective in prompting vocal characteristics that are very similar to the ones used in the classroom. Teachers gave the same lesson in their classrooms and later in front of a VR audience. Results showed that, in line with Niebuhr and Michalsky (2018), performing both in front of real and virtual audiences (compared to free speech performed before the experimenter in a control condition) significantly increased the participants' f0 values, their f0 variations and their voice intensity levels. Another recent study by Valls-Ratés et al. (2021) utilizing the same corpus used in the present study, with 31 participants, found that VR trainings induced a more audience-oriented prosody, making participants increase their f0 values, they spoke for longer time, there was an increase

in the number of pauses, and they also increased their gesture rate throughout the VR sessions. A study by [Notaro et al. \(2021\)](#) analyzed the effects of VR on fluency and gesture rate after 13 participants (20–25 years old) performed the same speech at two different times: the first time in front of a real audience and the second time in front of a VR audience, while also having the same real audience in front of them. They analyzed vocal parameters during VR and audience-based training and concluded that participants had a higher voice modulation, more voice power and paused more often when using VR. They also lowered their speech rate as well as their number of gestures per minute, pointing to the possibility that there existed a higher control over gestures while speaking with the VR glasses on. Finally, focusing on an L2 setting, [Thrasher \(2022\)](#) conducted a study with 25 participants (22 years old, L2 learners of French) that lasted 9 weeks. In order to assess the L2 speech in VR and Non-VR contexts, participants were asked to perform four public speaking tasks, two VR tasks and two in-class tasks. When French raters assessed the audio files, they found that the speech of participants using VR was more comprehensible than the speech of participants performing in-class.

Given that the studies reported in this section have shown that using VR for public speaking tasks triggers a more listener-oriented speech style, it is plausible to expect that a VR-training paradigm will trigger a more audience-oriented speech style in post-training speaking tasks. Yet to our knowledge very few studies have assessed the effects of VR on public speaking performance (see the next section).

Effects of VR on public speaking performance after training

To our knowledge, only two studies have been conducted to assess the effectiveness of VR public speaking training on public speaking performance after training. In a recent study, [Sakib et al. \(2019\)](#) performed a 3-month VR public speaking training study with a pre- and post-test design with 26 participants. Pre- and post-training speeches were performed in front of a real audience, whereas treatment consisted of 8 sessions in front of VR audiences. They collected a variety of measures of self-assessed and physiological anxiety, as well as ratings on speech performance assessed by external raters using an assessment form to rank speaker's performance from 1 (highest score) to 5 (lowest score). Results showed that participants improved their public speaking performance from pre- to post-training and also significantly reduced their self-assessed anxiety indicators, as well as two physiological anxiety measures (skin conductance response and skin temperature), resulting in a match between self-assessed and physiological markers. Even though the study concluded that VR environments were effective in reducing speakers' anxiety and enhancing public speaking performance,

there was no control group to compare these results to and public speaking performance was assessed in general terms. The second between-subject study by [Van Ginkel et al. \(2020\)](#) compared general public-speaking performances before and after VR public speaking training by involving both a VR and a Non-VR control group. The authors conducted a VR training study with 22 pre-university students across a 2-week period that consisted of three sessions: in the first and third sessions participants were introduced to the different features that an effective speech should include and after the instruction they had to give a 5-min speech in front of their peers. The second session was dedicated to performing a 5-min speech within a virtual environment, after which in a follow-up third session the VR condition received computer-mediated automatic immediate feedback and the control condition received delayed feedback given by an expert. The authors concluded that the VR session together with the given feedback was effective in improving eye contact and pace when delivering a speech in front of a real audience. However, they also pointed out that it is difficult to claim that the results are a direct consequence of the VR practice itself, as the instructions given to them, the feedback, and the independent practice could have had an influence as well.

Interestingly, in an L2 language learning context, [Gao \(2022\)](#) conducted an 8-week public speaking training study in which 90 Chinese university students participated in either a VR condition or a control condition based on traditional multimedia technology to test their proficiency in spoken English. After 8 weeks of autonomous learning, students were tested at post-training with English reading materials and oral presentation of specific topics. While participants in both conditions were successful in improving the oral English pronunciation skills (in this study they add the role of speech emotion to the usual pronunciation assessment systems that consider only the tone, intonation and rhythm of speech), the VR condition outperformed the control condition.

All in all, the investigations assessing the value of public speaking VR training initially point out to a gain in public speaking performance in terms of general performance, eye gaze and speech rate ([Sakib et al., 2019](#); [Van Ginkel et al., 2020](#)). Importantly several studies have indicated that VR triggers a more listener-oriented speech style ([Niebuhr and Michalsky, 2018](#); [Niebuhr and Tegtmeier, 2019](#); [Notaro et al., 2021](#); [Remacle et al., 2021](#); [Valls-Ratés et al., 2021](#)). Yet to our knowledge no previous investigation has assessed the value of VR training by assessing public speaking performance at post-test by incorporating a full-fledged prosodic analysis of the post-test speeches. We expect that the observed effect of VR in triggering an audience-oriented speech style will also carry over into the speakers' post-training speeches.

The present study: Main goal and hypotheses

Against the outlined research background, still very little is known about the potential boosting effects of practicing oral presentations with VR on developing students' public speaking skills and whether the training has an impact on the prosodic and gestural characteristics of the post-test speeches. Therefore, the main goal of this study is to investigate, through a between-subjects training experiment, whether training in public speaking with VR environments makes a difference in the overall quality of the oral presentations that students perform in front of an audience after training. To our knowledge, this is the first VR public speaking training experiment conducted with high school students that investigates not only the effects of training with VR on self-perceived anxiety both in the pre- and a post-training public speaking tasks but also on overall public speaking performance (through the use of persuasiveness and charisma ratings), as well as on oral presentation quality through a systematic analysis of the prosodic and gestural features of those oral presentations. Importantly, the assessment of the two speeches given in front of a live audience, e.g., before and after training, will be comprehensive. First, we will assess how the speaker feels in terms of self-perceived anxiety. Second, we will also include assessments about the persuasiveness of the speakers' charisma by external raters that are blind to the conditions. In addition, we will assess the prosodic characteristics of these speeches (understood holistically as involving a set of parameters including f_0 , tempo and voice quality characteristics), as well as the gesture rate, and the level of participants' own satisfaction after the training.

The following hypotheses will be tested: (a) Compared to the Non-VR public speech training, VR-based speech training will help diminish public speaking anxiety in the post-training public speaking task in front of a real audience. (b) VR public speaking training will lead to higher persuasion and charisma ratings. (c) VR public speaking training results in prosodic differences compared to the baseline condition of speakers, making the resulting speech more audience-oriented. (d) The audience-oriented prosody will be associated with a higher number of gestures in the VR condition. (e) Participants of the VR condition find more enjoyment and report a higher motivation for their future oral presentations.

In sum, the purpose of this educational intervention was to examine the impact of VR public speaking training on the quality of public speeches performed after training in front of a live audience, by comparing it to a Non-VR condition in which speeches were rehearsed individually. An important component of this assessment includes a complete analysis of the prosodic features of these speeches. In this way, we assess the value of a complementary use of a VR tool that can help educators promote the rehearsal of oral presentations and ultimately improve students' oral skills.

Methods

We designed a between-subjects training experiment with a pre- and post-test experimental framework. The public speaking training involved three training sessions, one per week (three for the VR condition and three for the Non-VR condition). Both before and after the training, a public speaking task was performed individually in front of a real audience, see [Figure 1](#). The total duration of the experiment, from the pre-training to the post-training public speaking task was 5 weeks.

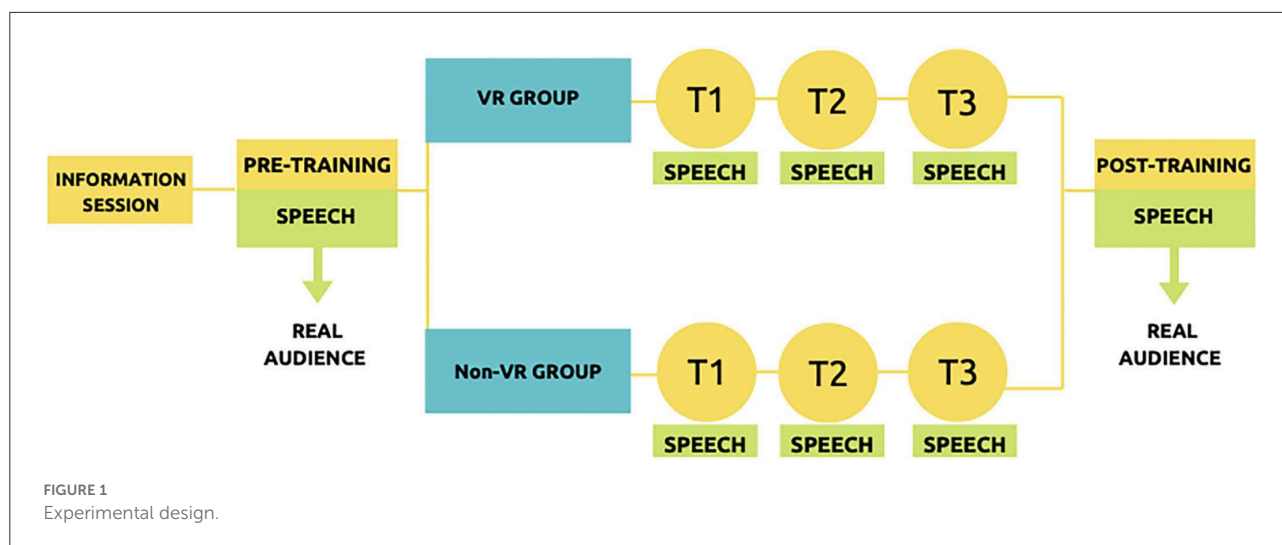
Participants

A total of 65 secondary school students aged 17–18 were recruited from four high schools (Institut Fort Pius, Institut Quatre Cantons, Institut Vila de Gràcia and Institut Icària) in the Barcelona area. These high schools are located in two central city quarters of Barcelona. The study was supported by the four school boards, which treated the proposed training as an extra-curricular activity which was carried out in the school premises. These four high-schools were chosen because they are placed in two central districts of Barcelona (Gràcia and Sant Martí), with very similar Catalan-Spanish language dominance (the percentage of Catalan speaking students being 81.9 and 78.8%, respectively), and with similar middle-income social composition.¹

Of the original 65 participants, 14 participants' data had to be disregarded for one of the following two reasons, namely (a) because of participants being absent at one of the training sessions or at the post-training phase, or (b) because their speeches at either pre- or post-test did not reach the minimum duration that we established (i.e., 1 min) or because they did not offer a minimum of two arguments to support their persuasive speech. The 50 remaining participants (mean age = 16.95, SD = 0.17; 70% female and 30% male) completed all five speeches with the required characteristics. Participants were randomly assigned to either the VR group ($N = 30$) or the Non-VR group ($N = 20$).

All participants were typically developing adolescents and had no history of speech, language, or hearing difficulties. Participation was voluntary, and all participants completed an informed consent form during the initial training session. Participants performed their speeches in Catalan. All students were bilingual Catalan-Spanish speakers, with 89.7% of them naming Catalan as their dominant language. The main language of instruction in the target schools is Catalan.

¹ Anuaris Estadístics de la Ciutat de Barcelona. 1996–2020 (*Barcelona's Statistical Annual Directory*): <https://ajuntament.barcelona.cat/estadistica/catala/Anuaris/Anuaris/anuari19/cap06/C0616010.htm>.



Materials for the public speaking tasks

A total of 5 short public speaking tasks had to be performed individually by each participant, two in front of a real audience (i.e., the pre-training and the post-training public speaking tasks), and three for training purposes. For all the public speaking tasks, participants were given a specific topic and a sheet of instructions (see Appendix) containing a list of arguments they could use in order to prepare a persuasive speech. In all cases, they were asked to prepare a 2-min speech.

An initial choice of 10 topics was first made based on a long list of suggested topics taken from a website maintained by instructors of public speaking and other communication courses (i.e., www.myspeechclass.com). This initial list of 10 topics was assessed through an online questionnaire which was distributed to mailing lists of 17-year-old boys and girls. A total of 58 anonymous students participated in the poll. They were asked to vote on their favorite topics from 1 (least liked) to 7 (most liked). The topic selected for both pre-training and post-training public speaking tasks was the same, namely: “Do you think that adolescents should spend more time in nature?”. In order to minimize the argumentation and expression differences across participants, five possible arguments were provided to participants. They were also given 2 min to prepare their speech. Though they could take notes for that purpose if they wished, they were not allowed to use the notes when they delivered their speech to prevent them from reading the whole speech.

The three topics for each of the three VR and Non-VR training sessions were the following: “What would the house of my dreams be like?”, “Is graffiti a form of art?”, and “Can happiness be bought?”. The instructions given to participants for the preparation of their speeches during the training sessions was the same as the instructions given to them for the pre- and post-test public speaking tasks.

Experimental design

The structure of this between-subjects training study was a pre-training phase followed by a training period and a post-training phase (see Figure 1). One week prior to the pre-training phase, an information session was organized by the experimenter in each of the high-schools and served the purpose of preparing the students for the pre-training session and explaining the experiment’s procedure and overall schedule that participants would have to bear in mind when delivering a speech. Pre and post-training sessions were also conducted by the experimenter and a research assistant. Both the research assistant and the 3-people live audience were blinded to the procedure of the study. During the information session participants were instructed on how to use VR and they could familiarize themselves with the VR goggles.

They were told that an audience of three people would attend their speech. They also knew that the pre-training speech would have to be persuasive, and that it was to be performed to convince three representatives of the Catalan Government to take action. Yet the topic itself would only be revealed to them immediately before the speech. After this, each group of students was randomly divided into the VR and the Non-VR group. The VR group performed the three training sessions delivering their speeches in front of a virtual audience, whereas the Non-VR group gave the same set of speeches while being alone in a classroom. The reason to choose three short VR sessions was based on the belief that adaptation to the virtual context would need some repetitions. Empirical reports of fast and reliable learning of visual context-target associations have proved effective after just three repetitions (Zellin et al., 2014). Finally, all participants carried out a post-training, which consisted of the same persuasive public speaking task as the pre-training.

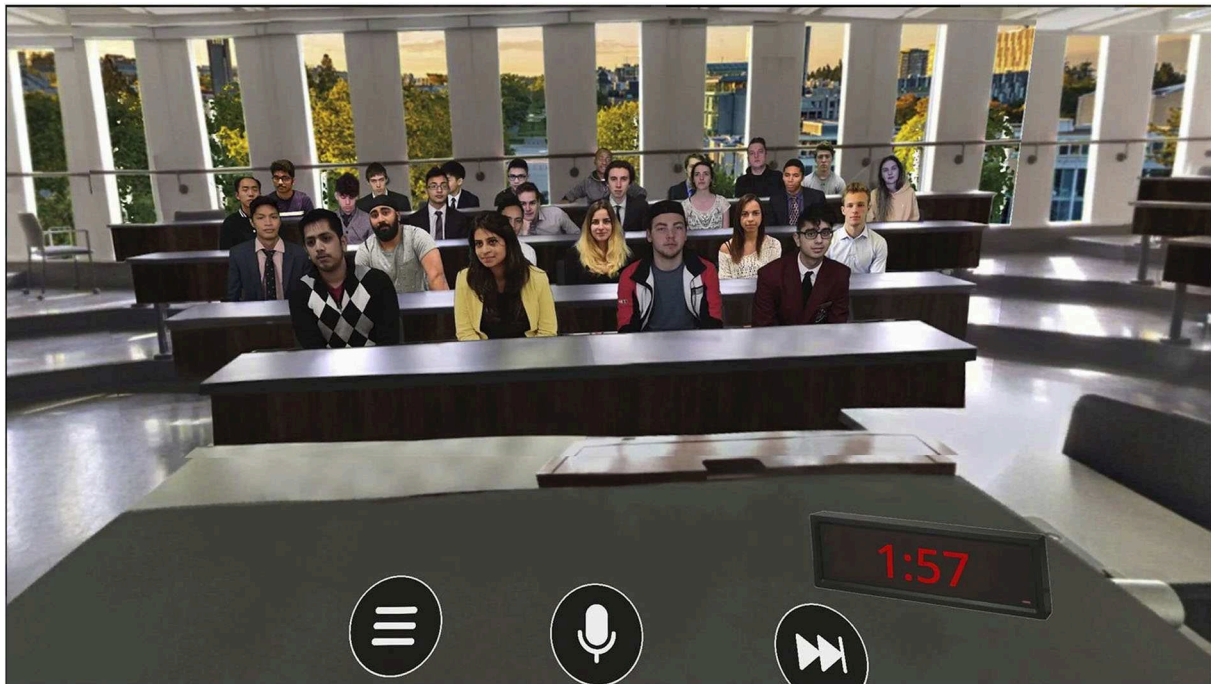


FIGURE 2
Screenshot of the VR scenario with a virtual audience generated by BeyondVR.

In order to pilot the materials, topics and procedure of the experiment, four 17-year-old students participated in a 3-h pilot session in which they were asked to prepare 3 speeches in 2 min to give in front of a small audience following our target set of instructions. The instructions informed participants of the amount of time they would have to prepare and to deliver the speech. For every speech they were given a written script of ideas related to the topic that they could use to include in their presentations. The pilot session contributed to refine and validate the final scripts and the procedure. For example, we realized that if speakers were allowed to use their written outline while speaking, they were reading from it most of the time. Therefore, we did not allow participants to have the outline with them to prevent them from reading and to enhance their connection with the audience.

Procedure

The experiment was performed individually in separate classrooms at the four high schools. The first author of the study was the experimenter and in charge of the data collection. All 5 public speaking tasks per student (3 during the training phase and 2 at pre- and post-training) were video recorded.

All participants started with the same pre-training task, which consisted of giving a brief speech in front of a live

audience. Before giving their speech, participants received a sheet of instructions in which they were asked to prepare and then deliver a 2-min persuasive speech in front of three representatives of the Catalan Department of Education to convince them to increase funding for secondary school field trips to the countryside. Participants were allotted 2 min to prepare their speech and did so alone in an empty classroom. After the 2 min of preparation had elapsed, they went to the adjacent classroom. The procedure was repeated for the post-training public speaking task.

For the training sessions, the procedure was largely similar between the two conditions. The Non-VR participants entered the classroom and were given the instructions. When they felt ready, they started performing the speech, with a visible timer that counted down the 2-min speaking time for them. For the VR participants, the only difference to the Non-VR participants was that right before practicing the speech, the experimenter fitted them with a Clip Sonic[®] VR headset to which a smartphone was attached. Using the free BeyondVR virtual reality interface application installed on the smartphone, the VR headset created the 3D illusion that the participant was standing in front of an audience. The virtual audience in this application moves while sitting and they show a sympathetic stance while the participant is speaking. They all look at the speaker and show interest in what the speaker is talking about, see [Figure 2](#). Note that a timer is also visible in the view provided

by the VR headset to allow speakers to monitor their use of time and not exceed the 2-min limit. Although we did not control for previous use of VR among participants, none reported any kind of discomfort wearing the VR goggles.

Anxiety measures

In order to control for anxiety and to facilitate comparisons with studies that have assessed anxiety in public speaking tasks through self-perception measures we used, as well as previous studies (e.g., Macinnis et al., 2010; Heuett and Heuett, 2011; Verano-Tacoronte and Bolívar-Cruz, 2015), the Subjective Units of Distress Scale, henceforth SUDS (Wolpe, 1990), a validated and widely used self-assessed anxiety questionnaire which uses a 100-point scale anchored on 0 (no fear), 25 (mild fear), 50 (moderate fear), 75 (severe fear), and 100 (very severe fear). Subjective distress refers to uncomfortable or painful emotions felt, and thus SUDS is used to systematically gauge the level of distress. The SUDS scale was developed by Wolpe (1969) and has been frequently used in Cognitive Behavioral Therapy (CBT) to evaluate treatment progress. Participants were given the SUDS assessment sheet just prior to entering the room where they would give their pre- and post-training speeches.

Satisfaction questionnaire

One month after the experiment ended, a brief online satisfaction questionnaire was sent to all participants asking the following three questions: “Did you feel comfortable participating in the experiment?”, “Did you have fun?” and “Did you find the experiment useful for your current oral presentations?”. They were asked to assess their satisfaction level using a Likert scale that ranged from 1 to 10. Nine (out of the 20) Non-VR participants and 19 (out of the 30) VR participants answered the online survey.

Data analysis

A total of 100 pre-training and post-training speeches were obtained from the 50 participants (50 participants \times 2 pre- and post-training speeches). The target persuasive speeches were assessed for the following features, namely (a) persuasiveness and charisma (Persuasiveness and charisma); (b) voice parameters (Voice parameters); (c) manual gesture rate (Manual gesture rate) and a satisfaction questionnaire (Satisfaction questionnaire). Apart from these measures on the actual speeches, a self-perceived anxiety SUDS measure and the results of the satisfaction questionnaire were also included in the data analysis.

TABLE 1 Survey questions regarding persuasiveness and charisma.

From 1 to 7 rate how persuasive the message is.	From 1 to 7 rate how charismatic the speaker is.
---	--

Persuasiveness and charisma

In order to assess the persuasiveness of pre- and post-training speeches, as well as the charismatic value of the speaker, a group of 15 raters (9 women and 6 men) with an age range from 23 to 63 years carried out a rating task on the speakers' persuasiveness and charisma, based on the video recordings of each presentation. The raters were chosen such that all had a university degree and that, overall, the rater sample was balanced with respect to gender. A 1-h training session was held with all raters and the first author of the study, in which they were given instructions as well as some time to practice and familiarize themselves with their task. They were first offered definitions of persuasiveness [understood by Rocklage et al. (2018, p. 751) as: “deliberate attempt to change the thoughts, feelings, or behavior of others”] and charisma [taking the definition by Niebuhr et al. (2020) “communication style signaling leadership qualities such as commitment, confidence, and competence that affect followers' beliefs and behaviors in terms of motivation, inspiration, and trust”]. Raters were asked to watch each video recording and then provide responses to the three questions in Table 1. They were asked to assess persuasiveness and charisma of the speaker in an intuitive way, without carefully analyzing vocabulary nor rhetorical strategies. They were asked to rate the speeches as if they were watching TV, assessing from 1 to 7 how persuasive the message was and how charismatic they perceived the speaker was.

An online survey sheet with the questions in Table 1 was prepared using Alchemer² (formerly SurveyGizmo, 2006). The 100 speeches were distributed across four surveys to offer the raters enough time to have a break after each block of about 15 stimuli. The speeches were presented in pairs. Each pair consisted of either pre- or post-training speeches of the same speaker so that raters could listen to them one after the other and assess which of the two was better. The rating task for all the speeches took about 5.5 h. The raters received a monetary compensation of 10 EUR per hour. The inter-reliability score (ICC) was excellent 0.913 (i.e., results are considered reliable as the score exceeded 0.7) (Koo and Li, 2016).

Voice parameters

For each participant, the total durations of the recorded speeches were similar in the pre- and post-training conditions ($M = 1:23$ min; $span = 1:00-2:00$ min). The acoustic analysis included a total of 16 different vocal parameters (5 f0 parameters,

² <https://www.alchemer.com/>

4 duration parameters, and 7 voice parameters; see below). The acoustic-phonetic analysis was automatically performed using the ProsodyPro script of Xu (2013) and the supplementary analysis script of De Jong and Wempe (2009), both with the (gender-specific) default settings of PRAAT (Boersma and Weenink, 2007).

In the f_0 domain, we measured f_0 minimum and maximum, the f_0 variability (in terms of the standard deviation), the mean f_0 and the f_0 range. For all five f_0 parameters, one value was determined per prosodic phrase. Measured values were checked manually for plausibility. Outliers or missing values were corrected by manual measurements. Moreover, all f_0 values were recalculated from Hz to semitones (st) relative to a base value of 100 Hz. The prosodic domain of calculation for those f_0 values was the interpausal unit (IPU), which was automatically detected. The criterion was the detection of an IPU boundary was the presence of a silent gap interval ≥ 200 ms, with silent gap being defined as a drop in intensity > 25 dB.

The tempo domain consisted of the following seven measured parameters: total number of syllables, total number of silent pauses (> 300 ms, which is above the perceived disfluency threshold in continuous speech) (Lövgren and Doorn, 2005), total time of the presentation (including silences), total speaking time (excluding silences), the speech rate (syllables per second including pauses), the net syllable rate (or articulation rate, i.e., syll/s excluding pauses) as well as ASD, i.e., the average syllable duration. ASD is a parameter that closely correlates with the fluency of speech (Rasipuram et al., 2016; Spring et al., 2019). As De Jong and Wempe (2009) summarize in their literature review: “An advantage of using inverse articulation rate [ASD] is that [...] it is a measure of disfluency, in the sense that higher values (longer mean syllable times) mean less fluent speech” (p. 900). All temporal measurements were conducted based on the analyzed presentation as a whole.

The domain of voice quality measurements included the nine parameters that are very frequently used in phonetic research (e.g., for analyzing emotional or expressive speech, see Banse and Scherer, 1996; Liu and Xu, 2014): harmonic-amplitude difference (f_0 corrected, i.e., $h1^* - h2^*$), cepstral peak prominence (CPP), harmonicity (HNR), $h1 - A3$, spectral center of gravity (CoG), formant dispersion (F1–F3), median pitch, jitter,³ and shimmer. Like for the f_0 parameters, voice-quality measurements were conducted based on the prosodic phrase, i.e., one value per prosodic phrase was calculated. Also, all values were manually checked and corrected, if required. This meant that a trained phonetician conducted a visual inspection of the measurement tables and marked potential outliers, i.e., in particular, unplausible values such as “0 Hz” or “600 Hz” for

mean f_0 and f_0 maximum or a F1–F3 formant dispersion of “–1 Hz”, etc. these were corrected by manual re-measurements (or deleted from the dataset).

Manual gesture rate

First, all communicative gestures were annotated by taking into account the gestural stroke (the most effortful part of the gesture that usually constitutes its semantic unit; (McNeill, 1992; Kendon, 2004)). Non-communicative body movements (self-adaptors, e.g., scratching, touching one’s hair; (Ekman and Friesen, 1969)) were excluded. Gesture rate was calculated per every speech as the number of gestures produced per speech relative to the phonation time in minutes (gestures/phonation time).

Satisfaction questionnaire

The means for each of the three questions of the satisfaction questionnaire and the reliability of the questionnaire (using Cronbach’s Alpha) were calculated.

Statistical analyses

The statistical analyses were performed using IBM SPSS Statistics 19. A set of GLMMs were run for five independent variables, namely SUDS (anxiety), Persuasion and Charisma, Voice and Gesture rate. The models include Condition (two levels: VR and Non-VR) and Time (two levels: Time 1-pre-training; Time 2-post-training) and their interactions as fixed factors. Subject was set as a random factor. Pairwise comparisons and *post-hoc* tests were carried out for the significant main effects and interactions.

For the satisfaction results, an independent a *t*-test was performed for each of the three questions in the satisfaction questionnaire. To make sure that there was rater interreliability, we performed a Reliability Analysis using the Intraclass Correlation Coefficient (ICC).

Results

Self-assessed anxiety SUDS

The GLMM analysis for SUDS showed a main effect of Condition [$F_{(1, 96)} = 8.785, p = 0.004$], which indicated that in general (both at pre- and post-training) Non-VR values were higher than VR values ($\beta = 13.792, SE = 4.653, p = 0.004$), and a main effect of Time [$F_{(1, 96)} = 10.807, p = 0.001$], showing that SUDS values were lower at post-training regardless of the condition ($\beta = 8.292, SE = 2.522, p = 0.001$). No significant interaction between Condition and Time was obtained, showing that the two conditions were not significantly

³ “The term jitter describes the small period-to-period variation in f_0 and hence deviation of a speaker’s voice from strict periodicity” (Niebuhr et al., 2020, p. 13).

TABLE 2 Summary of the GLMM analyses for the 5 f0 variables, in terms of main effects and interactions.

Variable	Main effect of time	Main effect of condition	Interaction time*condition
f0 min	$F_{(1, 94)} = 0.112, p = 0.738$	$F_{(1, 94)} = 7.171, p = 0.009$	$F_{(1, 94)} = 0.663, p = 0.417$
f0 max	$F_{(1, 94)} = 0.351, p = 0.555$	$F_{(1, 94)} = 14.073, p < 0.001$	$F_{(1, 94)} = 0.032, p = 0.859$
f0 variability	$F_{(1, 92)} = 4.155, p = 0.044$	$F_{(1, 92)} = 1.784, p = 0.185$	$F_{(1, 92)} = 3.329, p = 0.071$
f0 mean	$F_{(1, 94)} = 0.036, p = 0.849$	$F_{(1, 94)} = 12.643, p < 0.001$	$F_{(1, 94)} = 1.270, p = 0.263$
f0 range	$F_{(1, 88)} = 0.625, p = 0.431$	$F_{(1, 88)} = 1.204, p = 0.276$	$F_{(1, 88)} = 4.838, p = 0.030$

TABLE 3 Summary of the GLMM analyses for the 3 duration variables, in terms of main effects and interactions.

Variable	Main effect of time	Main effect of condition	Interaction time*condition
Number of syllables	$F_{(1, 88)} = 2.171, p = 0.144$	$F_{(1, 88)} = 0.662, p = 0.418$	$F_{(1, 88)} = 0.013, p = 0.911$
Number of silent pauses	$F_{(1, 88)} = 0.303, p = 0.584$	$F_{(1, 88)} = 1.347, p = 0.249$	$F_{(1, 88)} = 0.009, p = 0.926$
Total time of the presentation	$F_{(1, 88)} = 0.927, p = 0.338$	$F_{(1, 88)} = 0.846, p = 0.360$	$F_{(1, 88)} = 0.847, p = 0.360$
Total speaking time	$F_{(1, 88)} = 2.530, p = 0.115$	$F_{(1, 88)} = 0.058, p = 0.811$	$F_{(1, 88)} = 0.731, p = 0.395$
Speech rate	$F_{(1, 94)} = 1.884, p = 0.173$	$F_{(1, 94)} = 4.020, p = 0.048$	$F_{(1, 94)} = 0.918, p = 0.340$
Net syllable rate	$F_{(1, 93)} = 0.743, p = 0.391$	$F_{(1, 93)} = 10.502, p = 0.002$	$F_{(1, 93)} = 5.676, p = 0.019$
ASD	$F_{(1, 93)} = 0.856, p = 0.357$	$F_{(1, 93)} = 7.260, p = 0.008$	$F_{(1, 93)} = 4.472, p = 0.037$

different in triggering SUDS differences in the post-training public speaking task.

Persuasiveness and charisma

The GLMM analysis for persuasiveness showed a main effect of Condition [$F_{(1, 88)} = 7.461, p = 0.008$], which indicated that Non-VR values were higher than VR values ($\beta = 9.869, SE = 3.613, p = 0.008$), revealing an imbalance in the values at pre-test across groups in the form of an offset toward generally higher persuasiveness ratings in the Non-VR group as compared to the VR group (both at pre and post-test). The interactions between Time and Condition were not significant, meaning that the training conditions did not have a significantly different effect on the persuasiveness scores at post-training.

Regarding charisma, the GLMM analysis showed a main effect of Condition [$F_{(1, 88)} = 10.625, p = 0.002$], which indicated that in general (both at pre- and post-training), Non-VR values were higher than VR values ($\beta = 12.216, SE = 3.748, p = 0.002$). The analysis also showed a significant interaction between Time and Condition [$F_{(1, 88)} = 4.245, p = 0.042$], which indicated that both at pre-training and post-training the scores for Charisma of the Non-VR group were significantly higher than of the VR group: pre-training ($\beta = 13.821, SE = 3.802, p < 0.001$), post-training ($\beta = 10.611, SE = 3.854, p = 0.007$).

Prosodic parameters

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 2 shows a summary of those GLMM analyses in terms of main effects (Time and Condition), as well as interactions between Time and Condition. Summarizing, a main effect of Time was obtained only for f0 variability, meaning that the post-training values in both groups were higher than the pre-training values. A main effect of Condition was obtained for 3 variables (namely, f0 min, f0 max, and f0 mean), meaning that the participants in the VR group obtained higher f0 values, and larger f0 ranges across both pre- and post-training phases. A significant interaction was obtained for f0 range but no significant *post-hocs* reached significance.

Tempo domain

Regarding the tempo domain, a set of 7 GLMMs were applied to our target 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 3 shows a summary of those GLMM analyses in terms of main effects (Time and Condition), as well as interactions

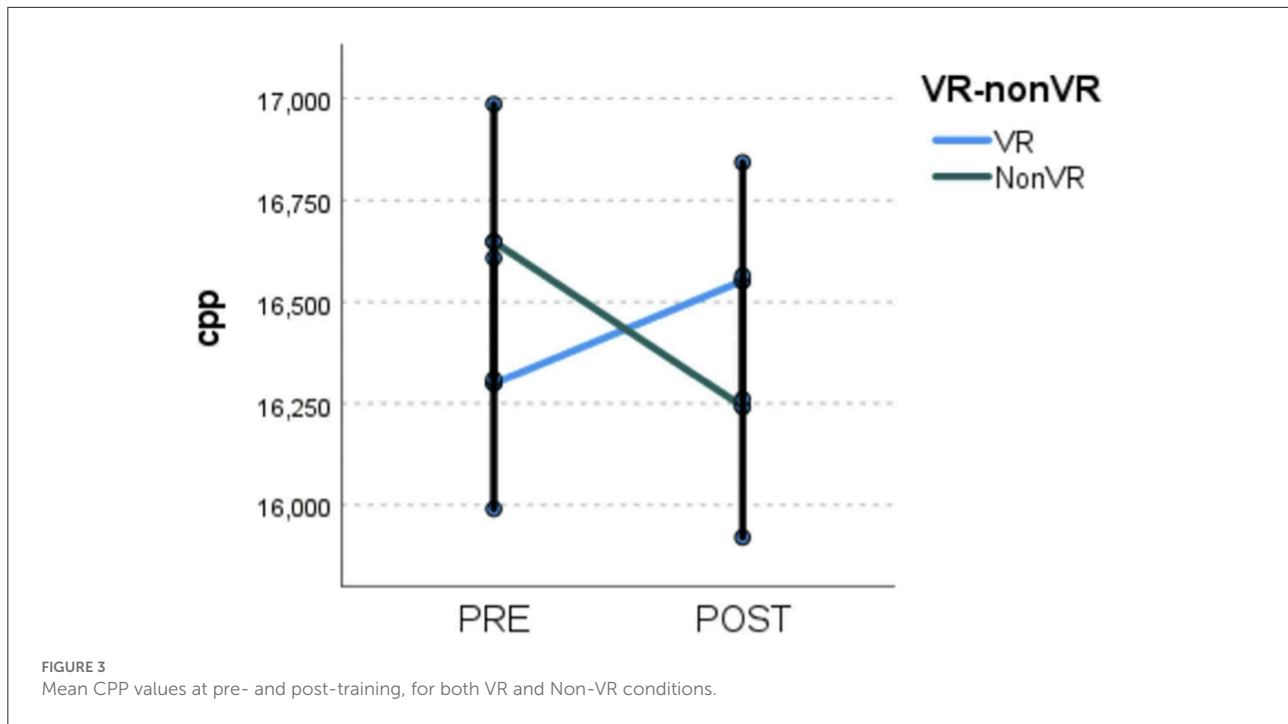


FIGURE 3
Mean CPP values at pre- and post-training, for both VR and Non-VR conditions.

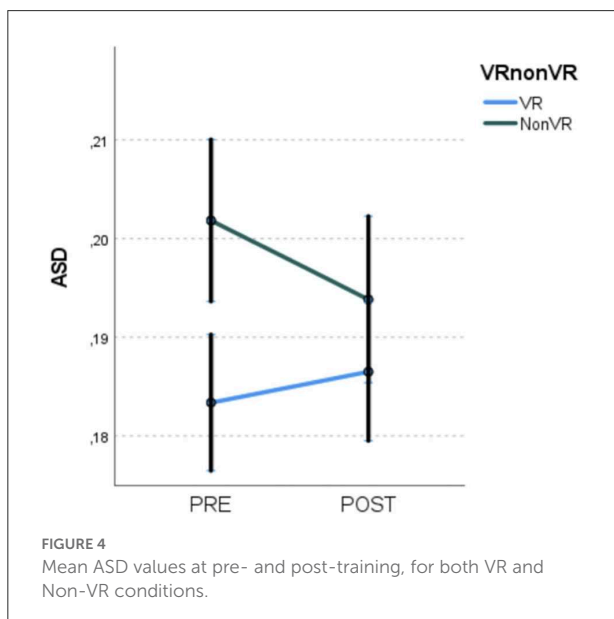


FIGURE 4
Mean ASD values at pre- and post-training, for both VR and Non-VR conditions.

between Time and Condition. Summarizing, no main effects of Time were obtained for any of the parameters of the duration domain. 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 obtained higher speech rate, net syllable rate (or articulation rate) values, and lower ASD values.

The variables that obtained significant interactions were net syllable rate and ASD. For net syllable rate (or articulation rate) in syl/s, the analysis revealed a significant interaction between Time and Condition [$F_{(1, 93)} = 5.676, p = 0.019$], which indicated that in the Non-VR group the values were significantly higher at post-training than at pre-training ($\beta = 0.211, SE = 0.099, p = 0.037$), while no significant differences were found in the VR group ($p = 0.241$). The interaction also showed that at pre-training there was a significant difference between the two groups, showing that the VR group values were higher than the Non-VR group values ($\beta = 0.544, SE = 0.143, p < 0.001$). With regard to ASD, the GLMM analysis showed a significant interaction between Time and Condition [$F_{(1, 93)} = 4.472, p = 0.037$], which indicated that in the Non-VR group the values were significantly lower at post-training than at pre-training ($\beta = 0.008, SE = 0.004, p = 0.050$), while no significant differences were found in the VR group ($p = 0.358$). VR-group speakers were thus able to maintain their lower ASD levels after training. The interaction also showed that at pre-training there was a significant difference between the two groups, showing that the VR group values were lower than the Non-VR group values ($\beta = 0.018, SE = 0.005, p = 0.001$). The GLMM analysis also showed a main effect of Condition [$F_{(1, 93)} = 7.260, p = 0.008$] which showed that VR values were lower than Non-VR values ($\beta = 0.013, SE = 0.005, p = 0.008$). Figures are provided in order to visualize the direction of the effects of the significant interactions. Figures 3, 4 show the mean syllable rate and ASD values obtained in the pre- and post-training tasks across conditions, respectively.

TABLE 4 Summary of the GLMM analyses for the 10 voice variables, in terms of main effects and interactions.

Variable	Main effect of time	Main effect of condition	Interaction time*condition
h1*-h2*	$F_{(1, 84)} = 0.168, p = 0.683$	$F_{(1, 84)} = 10.774, p = 0.002$	$F_{(1, 84)} = 0.215, p = 0.644$
h1-A3	$F_{(1, 84)} = 18.587, p < 0.001$	$F_{(1, 84)} = 8.523, p = 0.004$	$F_{(1, 84)} = 1.444, p = 0.233$
CPP	$F_{(1, 84)} = 0.334, p = 0.565$	$F_{(1, 84)} = 0.070, p = 0.792$	$F_{(1, 84)} = 17.009, p < 0.001$
Harmonicity	$F_{(1, 84)} = 2.197, p = 0.142$	$F_{(1, 84)} = 1.633, p = 0.205$	$F_{(1, 84)} = 0.091, p = 0.763$
CoG	$F_{(1, 84)} = 13.983, p < 0.001$	$F_{(1, 84)} = 0.004, p = 0.951$	$F_{(1, 84)} = 0.443, p = 0.507$
Formant dispersion 1–3	$F_{(1, 84)} = 9.417, p = 0.003$	$F_{(1, 84)} = 1.634, p = 0.205$	$F_{(1, 84)} = 0.023, p = 0.880$
Median pitch	$F_{(1, 82)} = 0.008, p = 0.930$	$F_{(1, 82)} = 14.938, p < 0.001$	$F_{(1, 82)} = 0.514, p = 0.475$
Shimmer	$F_{(1, 84)} = 0.104, p = 0.748$	$F_{(1, 84)} = 3.905, p = 0.051$	$F_{(1, 84)} = 4.195, p = 0.044$
Jitter	$F_{(1, 84)} = 1.845, p = 0.178$	$F_{(1, 84)} = 0.323, p = 0.571$	$F_{(1, 84)} = 3.677, p = 0.059$

Voice quality domain

In the domain of voice quality measurements, a set of 9 GLMMs were applied to our target variables, as explained in section Voice parameters above, namely h1*-h2*, h1-A3, CPP, HNR, CoG, formant dispersion, median pitch, shimmer, and jitter. Table 4 shows a summary of those GLMM analyses in terms of main effects (Time and Condition), as well as interactions between Time and Condition. A set of 9 GLMMs were applied to our target variables, namely h1*-h2*, h1-A3, CPP, HNR, CoG, formant dispersion, median pitch, shimmer, and jitter. Summarizing, a main effect of Time was obtained for 4 variables, namely h1*-h2*, h1-A3, CoG and formant dispersion, meaning that pre-training values were lower at pre-training across groups. A main effect of Condition was obtained for 4 variables, namely h1*-h2*, h1-A3, median pitch and shimmer, meaning that the participants in the VR group obtained higher values compared to the Non-VR group, both at pre and post-trainings.

Significant interactions were obtained for two variables, namely CPP and shimmer and a nearly significant interaction for jitter: For CPP, the GLMM analysis showed a significant interaction between Time and Condition [$F_{(1, 84)} = 17.009, p < 0.001$], which indicated that in the Non-VR group the values were significantly lower at post-training than at pre-training ($\beta = 0.351, SE = 0.112, p = 0.002$), and significantly higher at post-training for the VR group ($p = 0.009$). Regarding shimmer, the GLMM analysis also showed a significant interaction between Time and Condition [$F_{(1, 84)} = 4.195, p = 0.044$], which indicated that at pre-test groups were significantly different ($\beta = 0.018, SE = 0.008, p = 0.039$). The GLMM analysis for jitter showed a near significant interaction between Time and Condition [$F_{(1, 84)} = 3.677, p = 0.059$], which indicated that Non-VR values were significantly higher at post-training ($\beta = 0.006, SE = 0.003, p = 0.035$). Figure 5 shows the mean CPP values obtained in the pre- and post-training tasks across conditions.

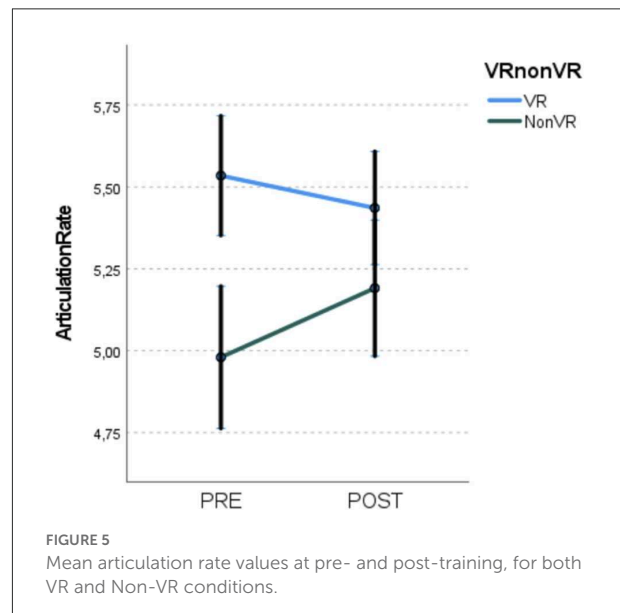


FIGURE 5 Mean articulation rate values at pre- and post-training, for both VR and Non-VR conditions.

Manual gesture rate

The GLMM analysis showed a significant interaction between Time and Condition [$F_{(1, 88)} = 4.796, p = 0.031$], but *post-hocs* did not reach significance. No main effects of Time and Condition were found.

Satisfaction questionnaire

Table 5 shows the descriptive results for the 3 questions in the satisfaction questionnaire, separated into VR and Non-VR conditions, on a scale from 1 to 10. As we can see, the responses to the latter two questions yielded higher ratings for the VR group than for the Non-VR group. Specifically, participants of the VR group had on average 0.33 scale points more fun with

TABLE 5 Descriptive results of the satisfaction questionnaire, separated into the VR and Non-VR conditions.

Question	VR group	Non-VR group
Did you feel comfortable participating in the experiment? (1–10)	Mean 8.15	Mean 8.66
	SD 1.22	SD 1.31
	N 19	N 9
Did you have fun? (1–10)	Mean 8.21	Mean 7.88
	SD 0.94	SD 1.46
	N 19	N 9
Do you find the experiment useful for your current oral presentations? (1–10)	Mean 8.10	Mean 6.22
	SD 1.24	SD 2.33
	N 19	N 9

the training task than their Non-VR counterparts and even considered that the perceived usefulness of the VR training was 1.88 scale points higher than their Non-VR counterparts. Yet while the latter difference is statistically significant [$t_{(28)} = 2.891$, $p = 0.004$], the other two are not. We also assessed the reliability of the questionnaire using Cronbach's Alpha. As the number of questions is <10 , it is considered that a good reliability score is ≥ 0.5 , and the Cronbach's Alpha score obtained was 0.725.

Discussion

The purpose of this experiment was to examine the impact of a 3-session VR public speaking training on the quality of the oral presentations of a group of 50 secondary school participants when speaking in front of a live audience. Specifically, we assessed the value of two complementary ways of rehearsing speeches, namely rehearsing with a VR audience or rehearsing alone in a room. To achieve this goal, we designed a between-subjects experiment with a pre-training, three training sessions and a post-training so that we could compare pre- to post-training speeches between a VR test condition and a baseline condition of Non-VR training. The duration between pre- and post-training was 5 weeks. One of the key contributions of this study is that it included a comprehensive assessment of the public speaking performance at pre- and post-trainings, specifically by assessing whether presenters in the post-training oral presentation achieved lower levels of anxiety, higher levels of persuasiveness/charisma, and/or a more audience-oriented speech from the point of view of prosodic and gestural features.

First, our results showed that the 3 training sessions reduced the anxiety levels of both VR and Non-VR groups of students to equal degrees in their post-training public speaking task. These results go in line with previous studies where VR

trainings proved effective in reducing self-assessed PSA levels of participants, both in clinical settings (e.g., Wallach et al., 2009, 2011; Lister et al., 2010; Lister, 2016; Lindner et al., 2018; Yuen et al., 2019; Zacarin et al., 2019) and in educational settings (e.g., Harris et al., 2002; Heuett and Heuett, 2011; Verano-Tacoronte and Bolívar-Cruz, 2015; Nazligul et al., 2017; Stupar-Rutenfrans et al., 2017). However, this result is not consistent with the hypothesis related to the stronger reduction of self-perceived anxiety in the VR group, as no differences were obtained for the VR and the Non-VR groups. Probably the reason why no differences were found between groups was due to the significant difference at pre-training (a 17-point difference higher for VR) that prevented VR speakers to reduce their self-perceived anxiety to a larger extent.

Second, ratings on persuasiveness and charisma did not result in any significant differences from pre-training to post-training in any of the conditions. This outcome is not consistent with our second hypothesis. As we will discuss later, having obtained no changes in f_0 patterns across groups might be the reason behind our results, as greater intonation changes would lead to higher charismatic speech (e.g., (Touati, 1993; Bosker and Kösem, 2017; Niebuhr and Fischer, 2019)), which was not found at post-training for any of the conditions.

Third, with respect to the effects of VR on prosodic parameters, the duration results show that Non-VR speakers significantly raised the articulation rate, i.e., they spoke at a faster pace in the post-training task. A similar change in pace is characteristic of the difference between carefully articulated, and audience-oriented spontaneous speech on the one hand and more self-directed and sloppy read speech on the other (see Jessen, 2007 for the tempo difference between a text-reading exercise and a communicative, spontaneous-speaking task). For ASD, Non-VR participants significantly decreased their values, meaning that they increased their fluency at post-training, but even with this increase were not able to reach the high level of fluency that the VR group was able to maintain at post-training. Voice-quality results show how VR speakers increased their CPP levels from pre-training to post-training speeches. Higher CPP levels are an indication that speakers' voices got clearer and more resonant and confident after training. Importantly, while the VR speakers significantly increased their clarity and resonance, the Non-VR speakers' voices, by contrast, got significantly less clear, resonant, and confident. Very likely this is caused by a reduced vocal effort, i.e., by a softer, less loud voice, produced with lower subglottal pressure.

Thus, overall, our prosody-related results favor an interpretation in which the VR training prevents speakers from falling victim to what Niebuhr and Michalsky (2018) termed the "erosion effect" of repetitive training while, at the same time, it favors a more audience-oriented voice quality in the post-training speeches. The erosion effect caused by repetitive training made the Non-VR speakers' presentations faster and less audience-oriented and their voices less powerful.

This finding is consistent with Niebuhr and Michalsky (2018) who also found that, compared to a control group of speakers who practiced their presentations without VR support, those speakers who could practice with VR support were significantly better able to suppress any negative effects of repetitive rehearsing on their speech prosody—and even improved in some aspects of their speech prosody. Since the lower the *jitter* value the more harmonic, less trembling and creaky the voice is, which suggests that the speakers of the VR group developed at post-training a clearer, stronger and less “shaky” voice, as it was also found by Notaro et al. (2021). Four variables obtained a main effect of Time $h1^*-h2^*$, $h1-A3$, CoG and formant dispersion, meaning that values of both conditions were lower at pre-training. As for a main effect of condition $h1^*-h2^*$, $h1-A3$, and median pitch values were generally higher for the VR condition.

Regarding the duration results, Non-VR speakers significantly raised the articulation rate, i.e., they spoke at a faster pace in the post-training task. For ASD, Non-VR participants significantly decreased their values, meaning that they reduced their fluency at post-training. The Non-VR group thus showed talking faster ($>$ art. rate) and reduce syllable durations ($<$ ASD), probably as a function of rate and fewer pitch accents. All in all, this is in our coaching experience the typical constellation of a bored, uninterested, routine presentation—that does not aim to get a message across to an audience but only to put words into sound.

Surprisingly, our results showed no significant changes across groups on f_0 values, meaning that intonation patterns did not change due to VR. At first glance this is inconsistent with the results of Remacle et al. (2021) where teachers performed the same lesson in class and with a virtual audience, or with the results of Niebuhr and Michalsky (2018) where participants had to train persuasive investor pitches with and without a VR audience. The important difference to the present study is, however, that both Niebuhr and Michalsky (2018) and Remacle et al. (2021) analyzed the prosody that speakers showed *during* VR immersion and not after it. As we already highlighted in the Introduction, to our knowledge our experiment is the first to analyze what happens (prosodically) when speakers take off the VR glasses and speak again to a live audience. In fact, as we report in a recent paper on the characteristics of speech during VR public speaking sessions (Valls-Ratés et al., 2021), the prosodic changes that we found when speakers perform public speaking tasks using VR (and Non-VR) are largely consistent with both Niebuhr and Michalsky (2018) and Remacle et al. (2021). F_0 -related melodic changes can basically be learned through training, as it has been demonstrated by Niebuhr and Neitsch (2020), where the training condition (unlike in our VR condition) included an explicit visualization and color-coded real-time evaluation of speech melody.

Fourth, regarding the use of gesture from pre- to post-training speeches, we did not find significant differences in the post-training task across conditions. We expected to observe a higher rate of gestures as a consequence of the more audience-oriented prosody observed in the VR condition, because research shows that “prominent parts of gestures (or gesture ‘hits’) tend to align with prosodically prominent parts of speech or pitch accents” (Cravotta et al., 2019, p. 1; see also, Shattuck-Hufnagel et al., 2007; Adrian and Clark, 2011; Loehr, 2012; Esteve-Gibert and Prieto, 2013; Esteve-Gibert et al., 2017). Therefore, our hypothesis regarding an increase in gesture rate for the VR condition is not supported.

Finally, an important result of our investigation is that 17-year-old students found the VR public speaking training (even in its basic, unguided form) more valuable to face their upcoming oral projects than the comparable, traditional rehearsing method without VR. This is also in line with other previous investigations by Kryston et al. (2021), Vallade et al. (2020) and Rodero and Larrea (2022). Thus, promoting more realistic and meaningful ways of individually rehearsing oral skills may enhance the whole experience of delivering a speech with regular and high-quality practice providing a cost-effective practice for education (Merchant et al., 2014; Boetje and van Ginkel, 2021) and increasing students’ motivation (Buttussi and Chittaro, 2018; Parong and Mayer, 2018). As we mentioned before, dealing with a high number of students per class and the extensive course curricula makes it extremely difficult for teachers to dedicate hours to enhancing oral skills in-class. Therefore, adopting VR technology could be of great help to make students rehearse individually and encourage them to practice oral skills regularly so as to become more confident and self-aware of their communicative strengths (Merchant et al., 2014; Van Ginkel et al., 2019) and acquire a more charismatic speech (Niebuhr and Michalsky, 2018; Niebuhr and Tegtmeier, 2019) in front of live audiences.

In summary, our study highlights the boosting effects of VR in terms of a handful of duration and voice quality parameters. In general, even though VR leads to preventing the erosion effect and to the use of a more clear and resonant voice after training, we need to acknowledge that this gain in audience-oriented prosody and public speaking confidence that the VR technology achieves, probably based on the presence effect (Slater and Sanchez-Vives, 2003; see section A complementary solution: Empirical evidence on the effects of VR for boosting public speaking skills), was not enough to obtain positive results in many of the other variables that were analyzed within prosodic parameters when the VR-trained speakers were in front of a live audience.

Moreover, a lower SUDS and a more clear voice quality achieved by the VR group were not enough to boost

persuasiveness and charisma scores after the training sessions. Therefore, the match that we expected to see between a more charismatic style in terms of prosodic parameters and the ratings on persuasiveness and charisma was not obtained and we can conclude that the changes in prosodic cues triggered by the VR training were not sufficient to promote a gain in those ratings.

The present study has some limitations. First, the study would have benefitted from a larger sample, which could have yielded more robust results and, thus, a clearer picture of how VR training sessions affect 17-year-old's public-speaking abilities. Second, even though anxiety was controlled through the use of the SUDS scale, a self-assessed measure, adding more objective instruments like electrophysiological measures would allow us to obtain a more fine-grained picture of the anxiety assessment of our participants and compare them with the subjective assessments. Third, in relation to persuasiveness and charisma, raters intuitively assessed the persuasiveness of the message. Even though all speeches contained at least two arguments, we acknowledge that we did not analyze or control for the strength of the arguments nor the rhetorical strategies used by each of the participants (cf. the Charismatic Leadership Tactics of Antonakis et al., 2011), which might have had an influence on the ratings. Fourth, in order to obtain positive effects on charisma and persuasiveness, as well as on f_0 parameters, the study could have added more (or longer) training sessions, together with explicit feedback strategies. We believe that giving specific instructions or using feedback strategies to participants (like in Niebuhr and Neitsch, 2020), could change the results at post-training, as seen in other studies (Chollet et al., 2015; Van Ginkel et al., 2019). Future longitudinal studies could be carried out in order to control for the students' perception of enjoyment and usefulness while using VR to ascertain whether the strong value that they assign to VR would remain constant or it is a result of the technology novelty. All in all, designing longer training sessions, longer periods of training, and adding feedback strategies could be regarded as future aims both in research and in practice.

In conclusion, the results of this study serve as a good starting point to continue developing our knowledge about the relationship between VR public speaking practice in secondary school education, self-confidence and the expected improvement in the quality of oral presentations.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

This study was approved by the Commission Project Ethical Revision (Comissió Institucional de Revisió Ètica de Projectes CIREP-UPF) and Recercaixa Project [2017 ACUP 00249] ethical approval. Written informed consent obtained from each participant and/or their legal representative, as appropriate.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This work benefited from funding awarded by the Spanish Ministry of Economy and Competitiveness (PGC2018-097007-B-I00, PID2021-123823NB-I00) and the Generalitat de Catalunya (2017 SGR_971). We also acknowledge the support from the Recercaixa Project (Recercaixa 2017ACUP 00249) and the Department of Translation at Universitat Pompeu Fabra through a 1-year doctoral grant to ÌV-R.

Acknowledgments

Thank you, young girls and boys of the four high-schools, for believing in the experiment and being so respectful throughout the study. Also to the high-school boards and teachers for being so supportive of the project. Thank you to Florence Baills, Mariia Pronina, and Patrick Rohrer (members of the GrEP group) for your help during data collection. And also GrEP group colleagues Júlia Florit Pons, Xiaotong Xi and Yuan Zhang for your great help with statistics. Thank you, board members, for being so patient and motivated to be part of the experiment. Thanks to Gemma Balaguer Fort, Elisenda Bernal, Gemma Boleda, and Emma Rodero for contributing to our research as committee members of the MA thesis and Ph.D. research defense, your questions have been so valuable. Finally, a special thanks to the 15 raters that assessed the persuasiveness and charisma of all participants.

Conflict of interest

Author ON is CEO of the speech-technology company AllGoodSpeakers ApS. Please see <https://oliverniebuhr.com/conflict-of-interest.html> for further information.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Adrian, T., and Clark, R. (2011). Audience perceptions of charismatic and non-charismatic oratory: the case of management gurus. *Leadership Q.* 22, 22–32. doi: 10.1016/j.leaqua.2010.12.004
- Antonakis, J., Fenley, M., and Liechti, S. (2011). Can charisma be taught? Tests of two interventions. *Acad. Manage. Learn. Educ.* 10, 374–396. doi: 10.5465/amle.2010.0012
- Armell, K. C., and Ramachandran, V. S. (2003). Projecting sensations to external objects: evidence from skin conductance response. *Proc. R. Soc. London Ser. B* 270, 1499–1506. doi: 10.1098/rspb.2003.2364
- Aymerich-Franch, L., and Bailenson, J. (2014). “The use of doppelgangers in virtual reality to treat public speaking anxiety: a gender comparison,” in *Proceedings of the International Society for Presence Research Annual Conference* (Vienna).
- Bailey, E. (2018). A historical view of the pedagogy of public speaking. *Voice Speech Rev.* 13, 31–42. doi: 10.1080/23268263.2018.1537218
- Banse, R., and Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.* 70, 614. doi: 10.1037/0022-3514.70.3.614
- Beatty, M. J. (1998). Public speaking apprehension, decision-making errors in the selection of speech introduction strategies and adherence to strategy. *Commun. Educ.* 37, 297–311. doi: 10.1080/03634528809378731
- Beatty, M. J., and Behnke, R. R. (1991). Effects of public speaking trait anxiety and intensity of speaking task on heart rate during performance. *Hum. Commun. Res.* 18, 147–176. doi: 10.1111/j.1468-2958.1991.tb00542.x
- Beatty, M. J., and Clair, R. P. (1990). “Decision rule orientation and public speaking apprehension,” in *Communication, Cognition, and Anxiety*, ed M. Booth-Butterfield (Newbury Park: Sage), 105–116.
- Behnke, R. R., and Sawyer, C. R. (1999). Milestones of anticipatory public speaking anxiety. *Commun. Educ.* 48, 165–172. doi: 10.1080/03634529909379164
- Boersma, P., and Weenink, D. (2007). *PRAAT: Doing Phonetics by Computer (Version 5.2.34)*. [Computer Software].
- Boetje, J., and van Ginkel, S. (2021). The added benefit of an extra practice session in virtual reality on the development of presentation skills: a randomized control trial. *J. Computer Assisted Learn.* 37, 253–264. doi: 10.1111/jcal.12484
- Boettcher, J., Carlbring, P., Renneberg, B., and Berger, T. (2013). Internet-based interventions for social anxiety disorder - an overview. *Verhaltenstherapie* 23, 160–168. doi: 10.1159/000354747
- Bosker, H. R., and Kösem, A. (2017). “An entrained rhythm’s frequency, not phase, influences temporal sampling of speech,” in *Proceedings of Interspeech 2017* (Stockholm), 2416–2420. doi: 10.21437/Interspeech.2017-73
- Bratman, G., Daily, G., Levy, B., and Gross, J. (2015). The benefits of nature experience: Improved affect and cognition. *Landsc. Urban Plan.* 138, 41–50. doi: 10.1073/pnas.1510459112
- Brown, T., and Morrissey, L. (2004). The effectiveness of verbal self-guidance as a transfer of training intervention: its impact on presentation performance, self-efficacy and anxiety. *Innov. Educ. Teach. Int.* 41, 255–271. doi: 10.1080/14703290410001733302
- Buttussi, F., and Chittaro, L. (2018). Effects of different types of virtual reality display on presence and learning in a safety training scenario. *IEEE Trans. Vis. Comput. Graph.* 24, 1063–1076. doi: 10.1109/TVCG.2017.2653117
- Choi, C. W., Honeycutt, J. M., and Bodie, G. D. (2015). Effects of imagined interactions and rehearsal on speaking performance. *Commun. Educ.* 64, 25–44. doi: 10.1080/03634523.2014.978795
- Chollet, M., Wörtwein, T., Morency, L. P., Shapiro, A., and Scherer, S. (2015). “Exploring feedback strategies to improve public speaking: an interactive virtual audience framework,” in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (ACM)* (Osaka), 1143–1154.
- Cravotta, A., Busa, M. G., and Prieto, P. (2019). Effects of encouraging the use of gestures on speech. *J. Speech Lang. Hearing Res.* 62, 1–16. doi: 10.1044/2019_JSLHR-S-18-0493
- Dalgarno, B., and Lee, M. (2010). What are the learning affordances of 3-D virtual environments?. *Br. J. Educ. Technol.* 41, 10–32. doi: 10.1111/j.1467-8535.2009.01038.x
- Daly, J. A., Vangelisti, A. L., and Weber, D. J. (1995). Speech anxiety affects how people prepare speeches: a protocol analysis of the preparation processes of speakers. *Commun. Monogr.* 62, 383–397. doi: 10.1080/03637759509376368
- Daniels, M. M. (2021). “Usability assessment of virtual reality as a training tool for oral presentation,” in *IOP Conference Series: Materials Science and Engineering* (Sanya).
- Daniels, M. M., Palaoag, T., and 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* (Sanya), 803.
- De Jong, N. H., and Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behav. Res. Methods* 41, 385–390. doi: 10.3758/BRM.41.2.385
- Ekman, P., and 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
- Esteve-Gibert, N., Borràs-Comes, J., Asor, E., Swerts, M., and Prieto, P. (2017). The timing of head movements: the role of prosodic heads and edges. *J. Acoust. Soc. Am.* 141, 4727. doi: 10.1121/1.4986649
- Esteve-Gibert, N., and Prieto, P. (2013). Prosody signals the emergence of intentional communication in the first year of life: evidence from Catalan-babbling infants. *J. Child Lang.* 40, 919–944. doi: 10.1017/S0305000912000359
- Fox Cabane, O. (2013). *The Charisma Myth - How Anyone Can Master the Art and Science of Personal Magnetism*. New York, NY: Penguin.
- Frisby, B. N., Kaufmann, R., Vallade, J. I., Frey, T. K., and Martin, J. C. (2020). Using virtual reality for speech rehearsals: an innovative instructor approach to enhance student public speaking efficacy. *Basic Commun. Course Ann.* 32, Article 6. Available online at: <https://ecommons.udayton.edu/cgi/viewcontent.cgi?article=1587&context=bcca>
- Gao, D. (2022). Oral english training based on virtual reality technology. *Eng. Intelligent Syst.* 30, 49–54. Available online at: <https://website-eis.crlpublishing.com/index.php/eis/article/view/1647>
- Goberman, A. M., Hughes, S., and Haydock, T. (2011). Acoustic characteristics of public speaking: anxiety and practice effects. *Speech Commun.* 53, 867–876. doi: 10.1016/j.specom.2011.02.005
- Gross, J. J., and John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *J. Pers. Soc. Psychol.* 85, 348–362. doi: 10.1037/0022-3514.85.2.348
- Gruber, A., and Kaplan-Rakowski, R. (2020). “User experience of public speaking practice in virtual reality,” in *Cognitive and Affective Perspectives on Immersive Technology in Education*, ed R. Zheng (Hershey: IGI Global), 235–249.
- Harris, S. R., Kemmerling, R. L., and North, M. (2002). Brief virtual reality therapy for public speaking anxiety. *Cyberpsychol. Behav.* 5, 543–550. doi: 10.1089/109493102321018187
- Heuett, B. L., and Heuett, K. B. (2011). Virtual reality therapy: a means of reducing public speaking anxiety. *Int. J. Humanities Soc. Sci.* 1, 1–6. doi: 10.3390/jpm10010014
- Howard, M. C., and Gutworth, M. B. (2020). A meta-analysis of virtual reality training programs for social skill development. *Comput. Educ.* 144, 103707. doi: 10.1016/j.compedu.2019.103707
- Jessen, M. (2007). Forensic reference data on articulation rate in German. *Sci. Justice J. Forensic Sci. Soc.* 47, 50–67. doi: 10.1016/j.scijus.2007.03.003

- Kahlon, S., Lindner, P., and Nordgreen, T. (2019). Virtual reality exposure therapy for adolescents with fear of public speaking: a non-randomized feasibility and pilot study. *Child Adolesc. Psychiatry Ment. Health* 13, 47. doi: 10.1186/s13034-019-0307-y
- Kendon, A. (2004) *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Koo, T. K., and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* 15, 155–163. doi: 10.1016/j.jcm.2016.02.012
- Kryston, K., Goble, H., and Eden, A. (2021). Incorporating virtual reality training in an introductory public speaking course. *J. Commun. Pedagogy* 4, 133–151. doi: 10.31446/JCP.2021.1.13
- Lane, H. C., Hays, M. J., Core, M. G., and Auerbach, D. (2013). Learning intercultural communication skills with virtual humans: feedback and fidelity. *J. Educ. Psychol.* 105, 1026–1035. doi: 10.1037/a0031506
- LeFebvre, L. E., LeFebvre, L., and Allen, M. (2020). “Imagine all the people”: imagined interactions in virtual reality when public speaking. *Imagin. Cogn. Pers.* 40, 189–222. doi: 10.1177/0276236620938310
- Legault, J., and Zhao, J., and Chi, Y., and Chen, W., and Klippel, A., and Li, P. (2019). Immersive virtual reality as an effective tool for second language vocabulary learning. *Languages* 4, 13. doi: 10.3390/languages4010013
- Lindner, P., Dagöo, J., Hamilton, W., Miloff, A., Andersson, G., Schill, A., et al. (2020). Virtual reality exposure therapy for public speaking anxiety in routine care: a single-subject effectiveness trial. *Cogn. Behav. Ther.* 50, 67–87. doi: 10.1080/16506073.2020.1795240
- Lindner, P., Miloff, A., Fagnäs, S., Andersen, J., Sigeman, M., Andersson, G., et al. (2018). Therapist-led and self-led one-session virtual reality exposure therapy for public speaking anxiety with consumer hardware and software: a randomized controlled trial. *J. Anxiety Disord.* 61, 45–54. doi: 10.1016/j.janxdis.2018.07.003
- Lister, H. (2016). *The effect of virtual reality exposure on fear of public speaking using cloud-based software* (Thesis), University of New Brunswick, Fredericton, NB, Canada.
- Lister, H. A., Piercy, C. D., and Joordens, C. (2010). The effectiveness of 3-D video virtual reality for the treatment of fear of public speaking. *J. Cyber Ther. Rehabil.* 3, 375–381. Available online at: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-79960359620&partnerID=40&md5=a5fe8e42ae54ec8f46e1272ef96a3748>
- Liu, X., and Xu, Y. (2014). Body size projection and its relation to emotional speech-reading in Mandarin Chinese. *Proc. Speech Prosody* 2014, 974–977. doi: 10.21437/SpeechProsody.2014-184
- Loehr, D. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Lab. Phonol.* 3, 71–89. doi: 10.1515/lp-2012-0006
- Lövgren, T., and Doorn, J. V. (2005). “Influence of manipulation of short silent pause duration on speech fluency,” in *Proceedings of Disfluency in Spontaneous Speech Conference* (Aix-en-Provence), 123–126.
- Macinnis, C., Mackinnon, S., and Macintyre, P. (2010). The illusion of transparency and normative beliefs about anxiety during public speaking. *Curr. Res. Soc. Psychol.* 15, 42–52. Available online at: <http://www.uiowa.edu/~grp/proc/crisp/crisp.html>
- McCroskey, J. (1982). Oral communication apprehension: a reconceptualization. *Ann. Int. Commun. Assoc.* 6, 136–170. doi: 10.1080/23808985.1982.11678497
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. University of Chicago Press.
- Menzel, K. E., and Carrell, L. J. (1994). The relationship between preparation and performance in public speaking. *Commun. Educ.* 43, 17–26. doi: 10.1080/03634529409378958
- Merchant, Z., Goetz, E., Cifuentes, L., Keeney-Kennicutt, W., and Davis, T. (2014). Effectiveness of virtual reality-based instruction on students’ learning outcomes in K-12 and higher education: a meta-analysis. *Comput. Educ.* 70, 29–40. doi: 10.1016/j.compedu.2013.07.033
- Mikropoulos, T., and Natsis, A. (2011). Educational virtual environments: a ten-year review of empirical research (1999–2009). *Comput. Educ.* 56, 769–780. doi: 10.1016/j.compedu.2010.10.020
- Morreale, S. P., Osborn, M. M., and Pearson, J. C. (2000). Why communication is important: a rationale for the centrality of the study of communication. *J. Assoc. Commun. Administr.* 29, 125. Available online at: https://www.academia.edu/29493673/Why_Communication_is_Important_A_Rationale_for_the_Centrality_of_the_Study_of_Communication
- Nazligul, M., Yilmaz, M., Gulec, U., Gözcü, M., O’Connor, R., and Clarke, P. (2017). “Overcoming public speaking anxiety of software engineers using virtual reality exposure therapy,” in *24th European Conference on Systems, Software and Services Process Improvement (EuroSPI 2017)* (Ostrava), 191–202.
- Niebuhr, O. (2021). “Advancing higher-education practice by analyzing and training students’ vocal charisma: evidence from a Danish field study,” in *7th International Conference on Higher Education Advances (HEAd’21)* (Valencia), 743–751.
- Niebuhr, O., Brem, A., Michalsky, J., and Neitsch, J. (2020). What makes business speakers sound charismatic? A contrastive acoustic-melodic analysis of Steve Jobs and Mark Zuckerberg. *Cadernos de Linguística* 1, 1–40. doi: 10.25189/2675-4916.2020.v1.n1.id272
- Niebuhr, O., and Fischer, K. (2019). “Do not hesitate! - unless you do it shortly or nasally: how the phonetics of filled pauses determine their subjective frequency and perceived speaker performance,” in *Proceedings of Interspeech 2019* (Graz), 544–548. doi: 10.21437/Interspeech.2019-1194
- Niebuhr, O., and 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* [International Speech Communication Association (ISCA)] (Poznan), 309–313.
- Niebuhr, O., and Neitsch, J. (2020). “Digital rhetoric 2.0: how to train charismatic speaking with speech-melody visualization software,” in *Lecture Notes in Computer Science*, Vol. 12335, *Speech & Computer*, eds A. Karpov and R. Potapova (New York, NY: Springer Nature), 357–368.
- Niebuhr, O., and Tegmeier, S. (2019). “Virtual reality as a digital learning tool in entrepreneurship: how virtual environments help entrepreneurs give more charismatic investor pitches, FGF studies in small business and entrepreneurship,” in *Digital Entrepreneurship*, eds R. Baierl, J. Behrens, and A. Brem (Cham: Springer), 123–158.
- North, M. M., North, S. M., and Coble, J. R. (2015). Virtual reality therapy: an effective treatment for the fear of public speaking. *Int. J. Virtual Real.* 3, 1–6. doi: 10.20870/IJVR.1998.3.3.2625
- Notaro, A., Capraro, F., Pesavento, M., Milani, S., and Busà, M. G. (2021). “Effectiveness of VR immersive applications for public speaking enhancement,” in *Proceedings of IS&T International Symposium on Electronic Imaging: Image Quality and System Performance XVIII* (Springfield, MA), 294-1–294-7.
- Parong, J., and Mayer, R. E. (2018). Learning science in immersive virtual reality. *J. Educ. Psychol.* 110, 785–797. doi: 10.1037/edu0000241
- Pearson, J. C., Child, J. T., and Kahl, D. H. Jr. (2006). Preparation meeting opportunity: how do college students prepare for public speeches? *Commun. Q.* 54, 351–366. doi: 10.1080/01463370600878321
- Peeters, D. (2019). Virtual reality: a game-changing method for the language sciences. *Psychon. Bull. Rev.* 26, 894–900. doi: 10.3758/s13423-019-01571-3
- Rasipuram, S., Rao, S. P., and Jayagopi, D. B. (2016). “Automatic prediction of fluency in interface-based interviews,” in *2016 IEEE Annual India Conference (INDICON)* (Bangalore: IEEE), 1–6.
- Remacle, A., Bouchard, S., Etienne, A., Rivard, M., and 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 Real.* 25, 935–944. doi: 10.1007/s10055-020-00491-1
- Rocklage, M. D., Rucker, D. D., and Nordgren, L. F. (2018). Persuasion, emotion, and language: the intent to persuade transforms language via emotionality. *Psychol. Sci.* 29, 749–760. doi: 10.1177/0956797617744797
- Rodero, E., and Larrea, O. (2022). Virtual reality with distractors to overcome public speaking anxiety in university students. [Realidad virtual con distractores para superar el miedo a hablar en público en universitarios]. *Comunicar* 72, 87–99. doi: 10.3916/C72-2022-07
- Rogers, R., Wooley, J., Sherrick, B., David Bowman, N., and Oliver, M. B. (2017). Fun versus meaningful video game experiences: a qualitative analysis of user responses. *Comput. Game J.* 6, 63–79. doi: 10.1007/s40869-016-0029-9
- Rubin, R. B., Rubin, A. M., and Jordan, F. F. (1997). Effects of instruction on communication apprehension and communication competence. *Commun. Educ.* 46, 104–114. doi: 10.1080/03634529709379080
- Ruscilla, J. J. (2019). *Virtual Reality and Job Interviews*. Academic Minute. Available online at: <https://academicminute.org>
- Safir, M. P., Wallach, H. S., and Bar-Zvi, M. (2012). Virtual reality cognitive-behavior therapy for public speaking anxiety: one-year follow-up. *Behav. Modif.* 36, 235–246. doi: 10.1177/0145445511429999
- Sakib, M. N., Chaspari, T., and Behzadan, A. (2019). “Coupling virtual reality and physiological markers to improve public speaking performance,” in *19th International Conference on Construction Applications of Virtual Reality (CONVR2019)* (Bangkok), 171–180.

- Shattuck-Hufnagel, S., Yasinnik, Y., Veilleux, N., and Renwick, M. (2007). "A method for studying the time alignment of gestures and prosody in American English: 'Hits' and pitch accents in academic-lecture-style speech," in *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue*, Vol. 18, eds A. Esposito, M. Bratanic, E. Keller, and M. Marinaro (Amsterdam: IOS Press), 34–44.
- Slater, M., and Sanchez-Vives, M. V. (2003). Enhancing our lives with immersive virtual reality. *Front. Robot. AI* 3, 74. doi: 10.3389/frobt.2016.00074
- Smith, T. E., and Frymier, A. B. (2006). Get'real': does practicing speeches before an audience improve performance?. *Commun. Q.* 54, 111–125. doi: 10.1080/01463370500270538
- Spring, R., Kato, F., and Mori, C. (2019). Factors associated with improvement in oral fluency when using video-synchronous mediated communication with native speakers. *Foreign Lang. Ann.* 52, 87–100. doi: 10.1111/flan.12381
- Stupar-Rutenfrans, S., Ketelaars, L., and Van Gisbergen, M. S. (2017). Beat the fear of public speaking: mobile 360° video virtual reality exposure training in home environment reduces public speaking anxiety. *Cyberpsychol. Behav. Soc. Netw.* 20, 624–633. doi: 10.1089/cyber.2017.0174
- Takac, M., Collett, J., Blom, K. J., Conduit, R., Rehm, I., and De Foe, A. (2019). Public speaking anxiety decreases within repeated virtual reality training sessions. *PLoS ONE* 14, e0216288. doi: 10.1371/journal.pone.0216288
- Thrasher, T. (2022). The impact of virtual reality on L2 French learners' language anxiety and oral comprehensibility: an exploratory study. *CALICO J.* 39, 219–238. doi: 10.1558/cj.42198
- Touati, P. (1993) "Prosodic aspects of Political rhetoric," in *Proc. ESCA Workshop on Prosody* (Lund), 168–171.
- Tse, A. Y. (2012). Glossophobia in university students of Malaysia. *Int. J. Asian Soc. Sci.* 2, 2061–2073. Available online at: <https://archive.aessweb.com/index.php/5007/article/view/2374>
- Vallade, J. I., Kaufmann, R., Frisby, B. N., and Martin, J. C. (2020). Technology acceptance model: investigating students' intentions toward adoption of 360° videos for public speaking rehearsals. *Commun. Educ.* 70, 127–145. doi: 10.1080/03634523.2020.1791351
- Valls-Ratés, I., Niebuhr, O., and Prieto, P. (2021). "Effects of public speaking virtual reality trainings on prosodic and gestural features," in *Proceedings of 1st International Conference on Tone and Intonation (TAI)* (Sønderborg), 214–218. doi: 10.21437/SpeechProsody.2022-33
- Van Ginkel, S., Gulikers, J., Biemans, H., Noroozi, O., Roozen, M., Bos, T., et al. (2019). Fostering oral presentation competence through a virtual reality-based task for delivering feedback. *Comput. Educ.* 134, 78–97. doi: 10.1016/j.compedu.2019.02.006
- Van Ginkel, S., Ruiz, D., Mononen, A., Karaman, A. C., de Keijzer, A., and Sitthiworachart, J. (2020). The impact of computer-mediated immediate feedback on developing oral presentation skills: an exploratory study in virtual reality. *J. Computer Assisted Learn.* 36, 412–422. doi: 10.1111/jcal.12424
- Verano-Tacoronte, D., and Bolivar-Cruz, A. (2015). "La confianza para hablar en público entre los estudiantes universitarios," in *Conference: XXIX AEDEM Annual Meeting, June 2015* (Donostia).
- Wallach, H. S., Safir, M. P., and Bar-Zvi, M. (2009). Virtual reality cognitive behavior therapy for public speaking anxiety: a randomized clinical trial. *Behav. Modif.* 33, 314–338. doi: 10.1177/0145445509331926
- Wallach, H. S., Safir, M. P., and Bar-Zvi, M. (2011). Virtual reality exposure versus cognitive restructuring for treatment of public speaking anxiety: a pilot study. *ISR. J. Psychiatry Relat. Sci.* 48, 91–97.
- Wilson, L., and Fullwood, C. (2017). The effect of immersion and presence in a virtual reality public speaking task. *Annu. Rev. CyberTherapy Telemed.* 17, 211–213. Available online at: <http://hdl.handle.net/2436/622249>
- Wolpe, J. (1969). *The Practice of Behavior Therapy*, Pergamon General Psychology Series, 1. 4th Edn. Elmsford, NY: Pergamon Press.
- Wolpe, J. (1990). *Pergamon General Psychology Series, 1. The Practice of Behavior Therapy*, 4th Edn. 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), 7–10.
- Xu, Y., Park, H., and Baek, Y. (2011). A new approach toward digital storytelling: an activity focused on writing self-efficacy in a virtual learning environment. *Educ. Technol. Soc.* 14, 181–191. Available online at: http://www.ifets.info/journals/14_4/16.pdf
- Yuen, E. K., Goetter, E. M., Stasio, M. J., Ash, P., Mansour, B., McNally, E., et al. (2019). A pilot of acceptance and commitment therapy for public speaking anxiety delivered with group videoconferencing and virtual reality exposure. *J. Contextual Behav. Sci.* 12, 47–54. doi: 10.1016/j.jcbs.2019.01.006
- Zacarin, M. R. J., Borloti, E., and Haydu, V. B. (2019). Behavioral therapy and virtual reality exposure for public speaking anxiety. *Trends Psychol.* 27, 491–507. doi: 10.9788/TP2019.2-14
- Zellin, M., Mühlénen, A., and von Müller, H. J. (2014). Long-term adaptation to change in implicit contextual learning. *Psychon. Bull. Rev.* 21, 1073–1079. doi: 10.3758/s13423-013-0568-z

Appendix

Instructions for the PRE- and POST-training public speaking tasks

ENGLISH VERSION

PREPARATION TIME: 2 MINUTES

SPEECH DURATION: 2 MINUTES

Situation: Three representatives of the Education Department have come to your high-school to listen to the proposals of a group of students. They're thinking of assigning more budget to school trips.

Your claim is that adolescents need to **spend more time in nature** and not so many hours inside the city schools.

In order to argue in favor of your proposal, you have prepared a list of studies with data that will allow you to **convince the representatives** to assign more budget to this field.

- More than 50% of the population lives nowadays in urban areas. It is estimated that in 2050 the number will increase up to a 70% (Bratman et al., 2015).
- People surrounded of less trees suffer more stress and higher mortality rates.
- Be surrounded of nature reduces the stress hormone, blood pressure and sugar in the blood.
- Be surrounded of nature increases cardiovascular and metabolic health, concentration and memory.
- Strolling in the forest increases creativity, vitality and relaxation (Finnish Forest Research Institute).

GOOD LUCK!

Instructions for the TRAINING 1 public speaking task

ENGLISH VERSION

THE HOUSE OF MY DREAMS

PREPARATION TIME: 2 MINUTES

SPEECH DURATION: 2 MINUTES

Script that can help you prepare the structure and content of the message:

- Description of the house
 - Place
 - Why would it be like that?
 - What would be essential to be part of the house?
 - What would you do in such a house?
 - Would you live alone or would you like to share it with other people?
- GOOD LUCK!

Instructions for the TRAINING 2 public speaking task

ENGLISH VERSION

IS GRAFFITI ART?

PREPARATION TIME: 2 MINUTES

SPEECH DURATION: 2 MINUTES

Script that can help you prepare the structure and content of the message:

- Description of what is a graffiti
 - Where do we usually find them
 - Who makes them
 - Why are they important/necessary or the opposite
 - What makes you state that it is art or not and why
 - Use examples and personal experience
- GOOD LUCK!

Instructions for the TRAINING 3 public speaking task

ENGLISH VERSION

MONEY CAN NOT BUY HAPPINESS

PREPARATION TIME: 2 MINUTES

SPEECH DURATION: 2 MINUTES

Script that can help you prepare the structure and content of the message:

- How would you describe happiness?
 - What does money buy and what doesn't?
 - Richness/Poverty
 - What makes you state or negate the topic sentence
 - Use examples or experience that can illustrate feelings of happiness
- GOOD LUCK!