Encouraging participant embodiment during VR-assisted public speaking training improves persuasiveness and charisma and reduces anxiety in secondary school students

Ïo Valls-Ratés [1,2]*[‡], Oliver Niebuhr [2‡] and Pilar Prieto [1,3‡]

[1]Department of Translation and Language Sciences, Universitat Pompeu Fabra, Barcelona, Spain, [2]Centre for Industrial Electronics, University of Southern Denmark, Sonderborg, Denmark, [3]ICREA, Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia

Practicing public speaking to simulated audiences created in virtual reality environments is reported to be effective for reducing public speaking anxiety. However, little is known about whether this effect can be enhanced by encouraging the use of gestures during VR-assisted public speaking training. In the present study two groups of secondary schools underwent a three-session public speaking training program in which they delivered short speeches to VR-simulated audiences. One group was encouraged to "embody" their speeches through gesture while the other was given no instructions regarding the use of gesture. Before and after the training sessions participants underwent respectively a pre- and a post-training session, which consisted of delivering a similar short speech to a small live audience. At pre- and post-training sessions, participants' levels of anxiety were self-assessed, their speech performances were rated for persuasiveness and charisma by independent raters, and their verbal output was analyzed for prosodic features and gesture rate. Results showed that both groups significantly reduced their self-assessed anxiety between the pre- and post-training sessions. Persuasiveness and charisma ratings increased for both groups, but to a significantly greater extent in the gesture-using group. However, the prosodic and gestural features analyzed showed no significant differences across groups or from pre-to post-training speeches. Thus, our results seem to indicate that encouraging the use of gesture in VR-assisted public speaking practice can help students be more charismatic and their delivery more persuasive before presenting in front of a live audience.

KEYWORDS

public speaking, virtual reality, anxiety, persuasion, charisma, prosody, gesture, embodiment

1 Introduction

Apart from improving their public speaking skills (Boyce et al., 2007), giving secondary school students the opportunity to practice public speaking has been shown to improve their social skills (Morreale et al., 2000), self-confidence, and acceptance by their peers (Bailey, 2018), while lessening the risk that they will not engage in critical thinking during class

(Blume et al., 2010). Given these potential benefits, it is clear that schools should provide as many opportunities for public speaking practice as possible. However, given the large number of students that many have to manage and the extensive syllabus they are expected to cover, teachers are often reluctant to devote much class time to practicing public speaking (Schneider et al., 2017), which also requires teachers to ensure that the social climate in the classroom is sufficiently safe and positive (Adler, 1980) for anxious students to overcome their fear of speaking to an audience (Kougl, 1980). Finally, students themselves are reported to put most of their preparation effort into writing the script of what they will say, spending at most 5 minutes on practicing their oral delivery (see Pearson et al., 2006).

Virtual reality technology (henceforth VR) can be used as a supplementary tool for rehearsing oral presentations or speeches in the classroom by means of a VR headset that gives wearers the visual 3-D illusion that they are standing in front of an artificially generated audience. The effectiveness of this tool in preparing students for speaking before real audiences has been demonstrated by research, as we will see below. However, in the present study we will explore whether combining such VR-assisted training with "embodiment" in the sense of an encouraged use of gestures while speaking will make student speakers both less anxious and more effective in subsequent experiences speaking to a live audience than VR-assisted practice in public speaking alone. Note that the current study is part of a set of three studies investigating VR-effects on public-speaking performance and public-speaking anxiety. The first two studies focused on learning after VR-assisted training compared to non-VR-assisted training. The present study focuses on gestures by comparing two VR-assisted conditions. Thus, the series of studies look at public-speaking performance and public-speaking anxiety across a sequence of training conditions, from non-VR to VR to gesture-activated VR (see Valls-Ratés, 2023 to appear, for an overview of the three studies).

This paper is organized as follows. In Section 1 we will discuss the utility of VR to train public speaking (1.1), previous literature on the value of VR for reducing public speaking anxiety (1.2), and training public speaking performance (1.3), and the role of embodiment in oral communication (1.4). Our methods are described in Section 2 and our experimental results in Section 3. Finally, a discussion and conclusions are offered in Section 4.

1.1 Using VR to train public speaking

While VR technology is now widely utilized for recreational purposes (Peeters, 2019), VR-simulated environments are also increasingly used in education to promote active learning (Legault et al., 2019). VR can elicit the subjective illusion known as *presence*, the illusion of "being there" in the scenery that the VR technology recreates, even though the user consciously knows that the environment depicted is simulated (Armel and Ramachandran, 2003). VR users feel immersed in this virtual environment (Slater et al., 2006) and engage in it as active participants, to a much more intense degree than what they experience when they use a laptop or phone (Bowman and Hodges, 1999; Slater and Sanchez-Vives, 2016). VR simulated environments have shown to be an effective learning tool (Mikropoulos and Natsis, 2011), in part because they

stimulate student enthusiasm and motivation (Dalgarno & Lee, 2010), to the extent that students are reported to be keen to adopt VR technology for their own educational purposes or encourage its adoption by educational institutions (Vallade et al., 2020).

With regard to training for public speaking in particular, research has shown that the speaking style of VR users addressing a simulated audience tends to be more listener-oriented in terms of its prosodic characteristics. To our knowledge, five studies have compared the features of speech when it is delivered to a live audience with speech delivered to a VR-simulated audience, three of them focusing on prosody. In the first, Niebuhr and Michalsky (2018) showed that the prosody of 24 university student participants as they practiced giving a speech in front of a VR-simulated audience was more conversational and listener-oriented than the prosody of students practicing alone, without an audience. The VR-assisted speech was characterized by a higher fundamental frequency (f01[1]) level, a larger f0 range, and a slower speaking rate. Interestingly, the speech of students practicing alone underwent an increasing "prosodic erosion" effect whereby the more the students repeated their speeches, the progressively lower and narrower the speech melody of their delivery became; by contrast, the VR-assisted speakers exhibited much less of this effect (see also Niebuhr and Tegtmeier, 2019). Also, VR-assisted speakers spoke for a longer time, made fewer pauses and used a higher intensity level. In the second study, which was carried out with 30 female elementary school teachers, Remacle et al. (2021) demonstrated that a VR-simulated classroom was able to induce in teachers' speech vocal features that were very similar to those they used in the classroom. In line with the findings by Niebuhr and Michalsky (2018), the participants' f0 values, f0 variation and voice intensity levels were all much higher in speech delivered to a class, whether real or simulated, compared to unprepared speech delivered to the experimenter. A similar example is the study by Selck et al. (2022), which showed that speakers adjust the vocal effort of their speech according to how far away the interlocutor is. Selck et al. found a similar adjustment to the speaker-listener distance also in VR dialogues, especially when the effect of visual immersion was complemented with a 3D acoustic-ambiance immersion. In the third study related to prosody (Valls-Ratés, 2023, chapter II), found that as secondary school students practiced before VR-simulated audiences, their prosody became audience-oriented, that is, stronger, more effortful and louder, although they did not perform more gestures.

The remaining two studies focused not only on the prosody of VR users but also on other features. Notaro et al. (2021) explored the effects of VR on the fluency and gesture rate of 13 participants who performed the same speech twice, first in front of a live audience and then in front of a VR-simulated audience but also in the presence of

---

1 Fundamental frequency (f0) refers to the rate at which the vocal cords vibrate during speech or singing. It is commonly measured in hertz (Hz) or cycles per second (cps). The f0 of an individual is primarily determined by the length of their vocal cords, which is correlated with their overall body size. Typically, f0 values range from 80 to 450 Hz, with males generally having lower voices than females and children (Bäckström et al., 2003). The phonational range of an individual, which is the range of frequencies they can produce, tends to decrease with age.

the same live audience. The authors concluded that participants' speech displayed larger f0 variation and higher intensity levels when they addressed the virtual audience. In the VR condition speakers also paused more often and reduced their speech rate as well as the number of meaningless gestures per minute, pointing to the possibility that when speaking to a VR audience they exerted greater control over their gestures. Finally, focusing on an L2 setting, Thrasher (2022) conducted a study with 25 learners of French performing two VR tasks and two classroom tasks to assess the impact of VR on the students' anxiety and French comprehensibility. Native French-speaking raters assessed the audio files and found speeches performed while using VR to be more comprehensible than speeches performed in the classroom. They also concluded that VR made participants less anxious than in-class tasks and they rated low-anxiety participants as easier to understand than high-anxiety participants, regardless of the performance context.

Overall, research suggests that speakers using VR to address a simulated audience are willing to adopt a more engaging listener-oriented way of speaking. Therefore, it is reasonable to expect that practicing public speaking using VR technology has the potential to not only improve the public speaking performance of high-school students but also in the process reduce *public speaking anxiety* (henceforth PSA).

1.2 The effect of VR-assisted training on public speaking anxiety

In line with current educational practices in Western countries, secondary-level students are increasingly expected to stand in front of the class and deliver expository talks, with their classmates and teacher as audience. Unsurprisingly, some students are more comfortable being the sole focus of attention than others, and a certain proportion of the students in any class may experience what has been labeled PSA when asked to present in front of an audience. Physiologically, PSA is manifested by a wide range of symptoms such as increased heart and breathing rates, nausea, a dry mouth or sweating (Smith et al., 2005; Boyce et al., 2007; Tse, 2012), but the psychological reality of PSA has been amply documented through the use of self-reported measures of anxiety.

In the last few decades, a body of research has shown that VR technology is useful to reduce PSA in clinical contexts (e.g., Wallach et al., 2009; Lister et al., 2010; Wallach et al., 2011; Lister, 2016; Lindner et al., 2018; Yuen et al., 2019; Zacarin et al., 2019) as well as in educational settings (see Daniels et al., 2020 for a review). However, this technology has not been the only treatment for anxiety and other types of phobias such as fear of heights, arachnophobia or claustrophobia. In the field of psychology, treatments such as Cognitive Therapy or Cognitive Behavioral Therapy have been widely employed to help patients reduce or overcome PSA (e.g.,Anderson et al., 2005; Wallach et al., 2009).

To our knowledge, four studies focusing on the impact of VR-assisted public speaking practice on PSA have been carried out in university settings, generally by comparing participant self-reported levels of distress, communication competence, willingness to communicate, and/or physiological measures before and after training sessions, and all of them found that VR has a stronger impact in reducing PSA than other approaches (Heuett and Heuett, 2011; Boetje and Van Ginkel, 2020; LeFebvre et al., 2020; Rodero and Larrea, 2022). Heuett and Heuett (2011) compared VR to Visualization treatment (Ayres and Hopf, 1985) and reported that, although both groups were successful at diminishing anxiety, VR participants significantly increased their willingness to communicate and their self-perceived communication competence. Boetje and Van Ginkel (2020) concluded that rehearsing with VR two times after having received feedback reduces PSA and improves oral skills more effectively than rehearsing only once. Rodero and Larrea's (2022) study with 100 university students investigated the role of distractors (e.g., someone coughing in the audience or a member of the audience asking a question) in participants' public speaking performance and anxiety. Comparing the performance of students who had rehearsed their speeches in the VR environment with that of students in a control group who had rehearsed their speeches in front of an instructor, they concluded that those practicing with VR reduced their self-assessed and physiologically measured anxiety significantly more than the control group. The authors speculated that the use of distractors more closely simulates what the speakers can expect from a live audience, making them feel more prepared and self-confident and more able to concentrate. The study by LeFebvre et al. (2020), involving one group of 17 students, also reported significant changes in PSA from pre- to post-test with students using VR to train their oral skills. Their results suggest that VR minimizes the cognitive strain on speakers when they rehearse because, unlike when they practice alone, they are freed from having to imagine the scene and setting of the live audience they will ultimately have to face.

To the best of our knowledge, only two studies have explored the role that VR environments can play in reducing PSA in secondary school students. In the study by Kahlon et al. (2019), a group of 27 adolescents (aged 13–16) diagnosed with PSA underwent a single 90-min therapist-led session in which they performed various oral exercises using VR. Participant self-reports at one and three months after the session showed diminished PSA levels, although the lack of control or comparison groups made it impossible to clearly identify the sources underlying this decrease. In the other study, carried out by the authors of the present paper, Valls-Ratés et al. (2022) compared the public speaking performance of 50 students before and after they had practiced giving a 2-min speech, either in front of a VR audience or alone in a classroom. Students assessed their own anxiety levels before and after rehearsing, and 15 independent raters also rated participant performance for persuasiveness in pre- and post-training speeches, which were in addition analyzed for prosodic features as well as gesture rate. Though both groups significantly reduced their self-perceived anxiety at post-training and developed a more audience-oriented prosody, the raters detected no significant differences in the persuasiveness of delivery nor in the charisma of speakers in either group.

1.3 The effect of VR-assisted training on public speaking performance

Several studies have assessed the potential benefits of VR-assisted public speaking training for mitigating PSA and boosting

public speaking performance. However, the few studies exploring the latter line of research came up with mixed results.

Sakib et al. (2019) performed a VR-assisted experiment involving 26 university students that included eight practice sessions with a pre- and post-test design consisting of giving a short speech before a live audience. Results showed improvements in the quality of the performance and self-assessed anxiety indicators at post-test. Nonetheless, the experimental design lacked a control element, limiting the external validity of the study's findings.

Similarly, in a study involving two groups of 11 pre-university students each, Van Ginkel et al. (2020) compared the effect of practicing a speech either using VR or alone in front of an instructor. Immediately after speaking, both groups received feedback. The feedback offered to members of the first group was based on immediate feedback automatically produced by the VR system regarding the speaker's use of voice, eye contact, and posture and gestures during the speech, while the second group received delayed feedback based simply on the instructor's direct observations. The authors concluded that in the VR condition both the VR environment and the feedback the VR system provided were effective at increasing eye contact and speech rate when participants gave their final speech to classmates in the last session of the study. Nevertheless, Van Ginkel et al. acknowledged that it was difficult to claim that the outcomes were a direct result of the VR-assisted rehearsal itself because the instructions received by participants, feedback, and practice outside the workshop might also have affected the results.

For their part, Kryston et al. (2021) analyzed how the quality of speech delivery by 140 students and their PSA levels were affected by practicing a speech in a VR-simulated setting compared to not practicing at all. Results indicated that VR training sessions did not affect the PSA self-reported by students, but that VR-assisted practice yielded higher quality speech ratings than no practice.

In the context of L2 learning, Gao (2022) compared a VR condition to a traditional multimedia technology condition to boost the English pronunciation skills of 90 Chinese university students. Results showed that both conditions were successful in improving oral English skills, but the VR condition outperformed the control condition.

On the whole, previous findings regarding the value of VR-assisted training for public speaking seem to point to a gain in general public speaking performance. Nonetheless, more research is needed to assess the impact of VR in public speaking training, especially in secondary education, where studies are scarce (Kahlon et al., 2019; Valls-Ratés et al., 2022).

1.4 Embodiment in VR-assisted training and in public speaking

The term *embodiment* refers to the interaction between the physical activity of our bodies and the (technological) environment, implying a strong connection between mind and body (Kilteni et al., 2012). Within the embodied cognition paradigm, body and environment have been related to cognitive processes and embodiment has been shown to be grounded in physical perceptive and motor systems (e.g., Barsalou, 1999; Shapiro, 2014). In this paper, we use the term embodiment to refer to the

participants' strong activation of the body's meaningful movements during VR public speaking experiences. Even though embodiment is related to the well-known 'sense of presence' in VR research (many authors have pointed out the correlation between higher levels of sense of presence and body movement; see Slater et al., 1995; Slater et al., 1998; Bianchi-Berthouze et al., 2007), here we will focus on encouraging participants' embodiment. That is, even though we will not measure or directly systematically vary participants' sense of presence, it is reasonable to assume that a higher amount of body engagement in creating nonverbal meanings (together with the speaker's prosody) will be not just more natural and effective, but it will also stimulate a higher sense of presence, for reasons outlined below.

The connection between body movements and the ensemble of sensations felt when a person is interacting with a VR-simulated environment was explored in a study by Slater et al. (1998) in which the researchers assessed the sense of presence of participants interacting with VR environments. Participants were asked to walk through a VR forest and count the trees with unhealthy leaves. In one condition, the trees varied from short to tall while in the other they were consistently taller than normal eye level. Thus, in the first condition participants had to turn their heads around and up and down and if necessary bend down, while in the second such movements were unnecessary. The authors found that participants who made more body movements while performing the tasks reported a significantly higher sense of presence (see also Slater et al., 1995). In a similar vein, Bianchi-Berthouze et al. (2007) found that body movement not only increased the engagement of participants, but also played a role in the affective way in which participants got involved in the task, resulting in engagement scores being positively correlated with how much the participant moved (see also Pallavicini and Pepe, 2020 for a decrease of participants' anxiety and body movement while playing VR video games). This body engagement is one of the factors that influences the sense of presence reported by VR users (Sanchez-Vives and Slater, 2005).

Outside the area of VR, the term embodiment has been used in the context of oral discourse performance to refer to the gesturing movements characteristically made by speakers when they speak, in other words, the participation of the body in the delivery of spoken messages. In the last few decades much of the literature has paid particular attention to how body movements and co-speech gestures are linked to language and thought (e.g., McNeill, 1992), that is, the way speakers use their faces, hands, or other body parts helps them express their ideas and, ultimately, is a reflection of their thinking (Hostetter and Alibali, 2019). Various theories have arisen in this connection, such as the gestures-as-simulated-action framework (e.g., Hostetter and Alibali, 2004; 2008; see also Hostetter and Alibali, 2018 for a review; see also Kita, 2000; McNeill, 2005), all of them sharing the view that embodied knowledge is directly reflected in speech-accompanying gestures.

Crucially, in the present paper we hypothesize that the encouragement of body engagement and the use of co-speech gesturing during VR-assisted public speaking training can trigger an improvement in public speaking performance. Research has shown that actively moving the body and gesturing while speaking (and even prompting an interlocutor to do so) facilitates language and cognitive processing tasks, perhaps because it increases access to words and neural activation

(e.g., Krauss et al., 2000). Gesturing has been shown to help communicate spatial imagery (e.g., Alibali, 2005) and perform complex motor tasks (Feyereisen and Harvard, 1999). The visual-spatial imagery of gesturing also seems to help speakers package spatio-motor information into units that are compatible with speech (e.g., Kita, 2000). Gesturing while explaining a task is a predictor of how soon speakers will master the task (Church and Goldin-Meadow, 1986; Pine et al., 2004), and spontaneously gesturing while performing a task improves memory retention (Alibali and Goldin-Meadow, 1993; Cook and Goldin-Meadow, 2006). Even the form of the gesture is important: a study by Thomas and Lleras (2009) with participants trying to solve a problem while occasionally either swinging their arms or moving them in other ways demonstrated that the participants could solve the problem more easily when swinging their arms than when performing other arm movements. The authors concluded that specific movements seemed able to guide learners' higher order cognitive processing. Importantly for the present study, previous studies have also shown that the experience of physical movement can have a direct effect on diminishing anxiety, as well as clinical depression (Wang et al., 2014; Gunnell et al., 2016; Korczak et al., 2017; McMahon et al., 2017). Repeated, rhythmic gestures in the form of aerobic exercises are negatively correlated with trait anxiety and depression and positively related to both physical health and self-concept (e.g., McDonald and Hodgdon, 1991; Fox, 2000). All in all, the results of this line of research indicate that the physiological changes triggered by one or multiple sessions of physical activity have a direct and positive effect on cognitive functioning (see Donnelly et al., 2016 for a review).

On a related note, different types of embodiment in public discourse have a clear effect on the listeners' assessments of the speeches; for example, the specific style of gesturing used by the speaker can directly influence the audience's evaluations. Specifically, various studies have found that listeners find gesturing speakers more self-assured and skilled (Maricchiolo et al., 2009), warmer and more in control of their performance (Gnisci and Pace, 2014) and more pleasant (Kelly and Goldsmith, 2004) than speakers who do not gesture. Despite this, some recent studies suggest that while audiences favor a moderate amount of gesture by speakers, excessive gesturing is felt to diminish the effectiveness of delivery as much as little or no gesturing (e.g., Rodero, 2022; Rodero et al., 2022). Posture also sends a message: various studies have shown that open postures convey high power and closed postures low power (Darwin, 1872; Hall et al., 2005; Carney et al., 2010). Other research suggest that postures not only send messages to viewers but also reinforce feelings of either dominance or submission in those who apply them, which can also make public speakers feel more or less self-confident (Cuddy et al., 2012). People who adopt high power poses feel more powerful, positive, in control, optimistic about the future, and focused on their ambitions (e.g., Anderson and Galinsky, 2006; Burgmer and Englich, 2012). However, evidence for the effect of power postures on speakers' feelings is mixed (e.g., Ranehill et al., 2015; Davis et al., 2017; Latu et al., 2017), and many of the existing studies are underpowered.

In sum, it seems that encouraging the use of embodiment during VR-assisted public speaking training has the potential to help boost oral skills after intervention and reduce the public speaking anxiety of participants. Crucially, within a VR simulation context, it might

well be that actively moving the body has an enhancing effect on the sense of presence that users experience, as has been reported by the studies reviewed in this section.

1.5 The present study: goals and hypotheses

Despite the considerable research outlined above, relatively few of these studies have focused on how VR could be used to improve training in public speaking skills, for secondary school students in particular. In addition, to our knowledge there has been no research so far on whether VR-assisted training in public speaking will be more effective—in terms of not only a more effective speaker performance but also reduced PSA—if speakers are encouraged to embody their speech during VR training, that is, to accompany their verbal message with moderate amounts of appropriate gesturing. Previous studies have shown that the use of VR does not automatically stimulate a more frequent use of gestures (e.g., Selck et al., 2022; Valls-Ratés et al., 2023). Therefore, the present study will investigate whether VR-assisted public speaking training in which participants are explicitly instructed to actively move their body will diminish speaker PSA and boost their public speaking performance after intervention to a greater degree than the same training without any instructions to use embodiment. Importantly, the study will include a comprehensive assessment of the students' public speaking performance before and after their VR-assisted training sessions which will include the participants' self-perceived levels of anxiety, listeners' perception of persuasiveness and charisma, and an assessment of the prosodic and gestural features of the pre- and post-training speeches.

The fundamental research question of the study is whether VR-assisted training that encourages an embodied delivery will improve speaker effectiveness and reduce self-perceived anxiety. We hypothesize that such training will 1) diminish speaker anxiety, 2) make the delivery of participants more audience-oriented in terms of specific use of prosodic features and gesture rate, and c) make participants sound more charismatic and their messages more persuasive.

2 Method

2.1 Participants

A total of 78 students aged 16 to 17 were recruited from four secondary schools located in two central city districts of Barcelona. Although the city of Barcelona is characterized overall by a high percentage of Catalan-Spanish bilingualism, the degree to which one or the other language dominates in a particular neighborhood varies considerably. However, the schools chosen here were selected on the grounds that the bilingualism of their student bodies (as well as the middle-class socio-economic status of their families[2]) would have fairly uniform features (on average, students at all four schools

---

reported that they used Catalan roughly 80% of the time in their daily lives).

Of the original 78 participants, data from eight participants had to be disregarded for one or both of the following two reasons: the participant failed to attend one of the practice trainings or perform the post-training task; and 2) their speeches in the pre- or post-training task lasted less than a minute or contained less than two supporting arguments. The mean age of the 70 remaining participants (71.43% female/28.57% male) was 16.45 years (SD = 0.36). All participants were typically developing adolescents and had no history of speech, language, or hearing difficulties.

The study was formally endorsed by the governing boards of all four schools, which treated the proposed training sessions as an extra-curricular activity that was carried out on the school premises.

2.2 Materials for the public speaking tasks

Since the experiment involved asking students to individually perform a total of five public speaking tasks, two in front of a real audience constituting the pre-training and post-training, and three in front of VR-simulated audiences constituting the practice sessions, it was felt necessary to control for the topics on which participants would speak on each occasion by mandating the same topic for each participant. In order to select topics that would be of interest to adolescents, an initial selection of 10 topics was made by the authors based on a long list of suggested topics taken from a public website for teachers of public speaking (www.myspeechclass. com). This list was fitted into an anonymous online survey asking respondents to rate on a seven-point scale how interesting they felt each topic would be, and a link to the survey was emailed to lists of about 75 17-year-olds, 58 of whom responded. The four topics receiving the highest scores overall from these respondents were chosen for the experiment.

For every speaking task, participants were provided with a set of printed instructions that included the topic for their speech and a list of five arguments they could employ to defend their ideas (see Supplementary Appendix). All participants received the same instructions. While the topic and arguments for the pre- and post-training speeches were identical, the topics for each of the three practice sessions were different, as were the accompanying arguments. Arguments provided were intended as guidance; participants were not required to use them in their speeches, nor were they told to employ a particular number of arguments.

The instructions and procedures of the experiment were piloted by four 17-year-old students in a 3-h session that enabled the researcher to refine and validate the final instructions and topics. The language of all materials and procedures was Catalan. It was also the language used by participants to deliver their speeches.

2.3 Experimental design

One week prior to the pre-training speech to a live audience, an information session was held by the experimenter in each of the high schools. The session served the purpose of explaining the experimental procedure and overall schedule. Participants were informed that the training period would consist of five sessions consisting of the preparation and delivery of a public speech, but that only the first and last sessions would be in front of a live audience, which would consist of three real people. Participants were also given the opportunity at this time to familiarize themselves with the use of VR goggles. Participants were specifically informed that their speeches had to be persuasive, since their audiences would consist of three representatives of the Catalan government who might be swayed to initiate policy (e.g., allocating more government spending to school field trips to the countryside) based on what they had heard.

After the information session, the researcher randomly divided participants from each school into two groups, both of which would participate in the subsequent public speaking practice sessions in front of a VR-simulated audience. One of the two groups, however, would be encouraged by the researcher to accompany their speech with gesture—henceforth the Gesture Activated VR group (n = 40) while the other would receive no instructions with regard to their use of gesture while speaking—henceforth the Non-Gesture Activated VR group (n = 30). Even though this study explores the differences in gesture encouragement while using VR, we considered that it was clearer to label the two groups "Gesture Activated VR" and "Non-Gesture Activated VR" group.

The rationale for planning three such sessions was that it was felt only one such session would provide insufficient time for the participant to become comfortable speaking in a VR-simulated environment. Research has shown that visual context-to-target associations can be learned effectively after three repetitions in VR (Zellin et al., 2014).

Though all participants performed the three practice speeches to a VR audience following the same basic instructions, the participants in the Gesture Activated VR group were given the following additional instruction in writing right before each of the three training sessions: "Remember to use your whole body to express yourself fully".

Finally, as noted above, all participants again performed a speech to a live audience of the same three "government representatives" as a post-training. The topic on which they were instructed to speak was identical to that used for the pre-training. The full duration of the experiment was 5 weeks. The experimental design is shown schematically in Figure 1.

2.4 Procedure

All public speaking performances were carried out individually by each participant in a silent room at each participating school and were video-recorded. They were supervised by the first author, who also managed the collection of data with the help of an assistant. For the pre- and post-training public speaking tasks, three 24-year-old university students also attended the session and acted as the live audience (the "government representatives"). Neither the research assistant nor the three members of the audience were aware of the goals of the study. To prevent our behavioral data from being biased by experimenter effects (see Rosenthal, 1976), the first author welcomed participants and informed them about the procedure but was present neither in the practice room nor in the room where participants gave the pre- and post-test speech.

Before the pre-training public speaking performance to the live audience, participants were given the written instructions and left alone for 2 minutes to mentally prepare what they

FIGURE 1
Experimental design.
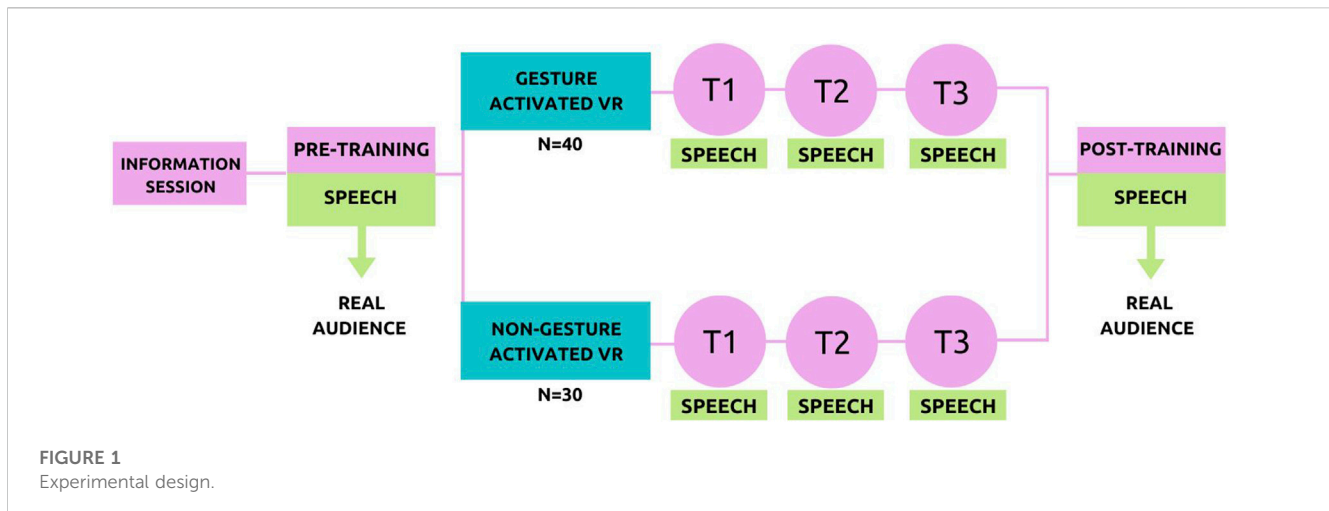
planned to say. The topic prompt was "Do you think that adolescents should spend more time in nature?" They then proceeded to the room where the "government representatives" were seated and delivered their speech. They were allowed a maximum of 2 minutes to do so.

The first of the three training sessions took place a week later, and the second and third were conducted over the following 2 weeks. As with the pre-training speech, participants had 2 minutes after receiving the written instructions to individually plan their speech. After the 2 minutes of preparation had elapsed, they went to the adjacent classroom, where the experimenter fitted them with a Clip Sonic® VR headset, to which a smartphone was attached. A week after the third training session, participants individually performed the post-training public speaking task, speaking about the same topic and to the same audience as in the pre-training task.

2.4.1 VR equipment

The study used a free-of-charge VR interface application installed on the smartphone called BeyondVR©. When the phone screen is viewed through special cardboard glasses, it gives the user the impression that they are standing in front of an audience of 40 people. You can find the screenshots of the virtual audience here. The computer-generated low-fidelity audiences make gestures and body movements resembling those that a live audience would make while listening to a speaker. However, the audiences generated by this application do not react to what the speaker says, nor can they be manipulated to behave in different ways. Participants were not able to see their own body while wearing the VR headset nor could they see a virtual representation of their body in the VR environment. Participants were able to monitor their speaking time by referring to a timer displayed in their field of vision by the headset.

2.5 Anxiety measures

Speaker anxiety was self-reported by participants just prior to entering the room where they would give their pre- and post-

training speeches using the Subjective Units of Distress Scale (SUDS; Wolpe, 1969). SUDS has been frequently used in cognitive-behavioral treatments and exposure practices to evaluate treatment progress, as well as for other research purposes. More specifically, the SUDS has been widely used in the analysis of speaker anxiety (e.g., North et al., 1998; Bartholomay and Houlihan, 2016; Takac et al., 2019) and is a validated instrument in which the reporting individual indicates his or her levels of anxiety in various contexts, using a 100-point scale where '0' represents no distress whatsoever and '100' represents the most intense distress imaginable. Each ten-point interval on the scale is accompanied by a brief description of how the participant might feel, so that the participant identifies with its meaning in the most specific way possible.

2.6 Public speaking performance measures

A total of 140 pre- and post-training test speeches were obtained from the 70 participants. They ranged from 1 to 2 min in duration, the mean being 1:23 min.

As noted above, these speeches were assessed for 1) perceived persuasiveness and charisma (2.6.1); 2) prosodic parameters (2.6.2); 3) and manual gesture rate (2.6.3).

2.6.1 Perceived persuasiveness and charisma

The impression created by each speech on a listener was measured in terms of the perceived persuasiveness of the speech and the perceived charisma of the speaker.

Persuasion has been defined as "the deliberate attempt to change thoughts, feelings, account, or behavior of others" (Rocklage et al., 2018: 1). More specifically, (Scheidel, 1967: 1) defines persuasion as "the activity in which the speaker and the listener are conjoined and in which the speaker consciously attempts to influence the behavior of the listener by transmitting audible and visual language". It has been shown that the perception of persuasion is modulated not only by the specific information transmitted by the speaker but also by the prosodic characteristics of the oral discourse (e.g., Burgoon et al., 1990; Krauss et al., 1996; Manusov and Patterson, 2006; Jackob et al., 2011; Yokoyama and Daibo, 2012), as well as by the gestural

performance (Mehrabian and Williams, 1969; Ekman et al., 1976; Kelly and Goldsmith, 2004; Maricchiolo et al., 2009; Peters and Hoetjes, 2017). For example, more varied intonation, greater fluency, and faster speaking rate are likely to convey more credibility and overall persuasiveness (Jackob et al., 2011), and greater vocal variety enhances the impression of competence, character, and sociability in a speaker (Addington, 1971; Ray, 1986).

Charisma has been widely studied, as it is a key aspect of leadership and social interaction. Contrary to the earliest definitions of charisma, which defined it as innate or almost magical (Weber, 1968), it is now regarded as an ability that can be taught and learnt. According to a recent terminological refinement of the concept by Michalsky and Niebuhr (2019), charisma represents a particular communication style. As (Niebuhr and Neitsch, 2020:358) point out, [charisma] gives a speaker leader qualities through symbolic, emotional, and value-based signals. Three classes of charisma effects are to be distinguished in the [public speaking] context, namely, 1) conveying emotional involvement and passion inspires listeners and stimulates their creativity; 2) conveying self-confidence triggers and strengthens the listeners' intrinsic motivation; 3) conveying competence creates confidence in the speakers' abilities and hence in the achievement of (shared) goals or visions. Inspiration, motivation, and trust together have a strongly persuasive impact by which charismatic speakers are able to influence their listeners' attitudes, opinions, and actions.

In the present study, a group of 15 raters (9 women and 6 men, aged 23 to 63, all university-educated) assessed speakers' persuasiveness and charisma based on the video recordings of the pre- and post-training test speeches. The first author of the study led a 1-h training session in which the raters, guided by the definitions of persuasiveness and charisma offered above, observed a public speaker and then rated their performance.

After training, the 15 raters were asked to watch each of the 140 video recordings embedded in an online questionnaire created using Alchemer (https://www.alchemer.com). After raters had viewed each speech, they were asked to answer two questions. "On a scale of 1–7, where 1 is "totally unpersuasive" and 7 is "extremely persuasive", rate the persuasiveness of the message" and "On a scale of 1–7, where 1 is "totally uncharismatic" and 7 is "extremely charismatic", rate the degree of charisma conveyed by the speaker" (see other studies that have employed perceptive ratings of charisma; e.g., Rosenberg and Hirschberg, 2009; Berger et al., 2017; Siegert and Niebuhr, 2021; Weninger et al., 2012). Raters were instructed to assess persuasiveness and charisma holistically and spontaneously, analyzing neither the words nor the rhetorical figures the speakers' employed. The scores for both persuasiveness and charisma variables ranged from 15 to 105.

The 140 speeches were presented in pairs in a randomized order to make it easier for raters to spot differences by comparing the same speaker at two different times. This was done to make ratings more sensitive, and while we increased sensitivity, we did not introduce a bias as the raters did not know that they were rating before–after comparisons. To avoid rater fatigue, the questionnaire was divided into several units. The assessment tasks for all presentations took approximately 6 hours in total. Raters received financial compensation of 10 euros per hour. The inter-reliability score (ICC) across raters was found to be excellent 0.904 (i.e., results

are considered reliable, as the score exceeded 0.7) (Koo and Li, 2016).

2.6.2 Prosodic measures

Acoustic-prosodic analysis of all 140 speeches was performed automatically by means of the ProsodyPro script by Xu (2013) and the supplementary analysis script by De Jong and Wempe (2009), both using the PRAAT (gender-specific) default settings (Boersma and Weenink, 2007). The analysis included a total of 20 different prosodic parameters, namely, five f0 parameters, seven duration parameters, and eight voice quality parameters.

The five f0 parameters were f0 minimum and maximum, f0 variability (in terms of the standard deviation), mean f0 and f0 range. A value was determined for each prosodic phrase for all five f0 parameters. Measured values were checked manually for plausability. Correction of outliers or missing values was performed by taking measurements manually. Additionally, all f0 values were recalculated from Hz to semitones (st) relative to a base value of 100 Hz. The prosodic domain of calculation for those f0 values was the interpausal unit (IPU), which was automatically detected. The criterion for the detection of an IPU boundary was the presence of a silent gap interval ≥300 ms, with silent gap being defined as a drop in intensity >25 dB.

The tempo domain consisted of the following seven 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 average syllable duration (ASD). ASD is a parameter that closely correlates with the fluency of speech (Rasipuram et al., 2016; Spring et al., 2019).

The domain of voice quality measurements included the eight 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 (f0 corrected, i.e., h1*-h2*), cepstral peak prominence (CPP), harmonicity (HNR), h1-A3, spectral center of gravity (CoG), formant dispersion (F1-F3), jitter, and shimmer. Voice quality measurements were based on the prosodic phrase, that is, one value per prosodic phrase was calculated. Also, all values were manually checked and, if necessary, corrected by a trained phonetician who conducted a visual inspection of the measurement tables and marked potential outliers, in particular, implausible values such as "0 Hz" or "600 Hz" for mean f0 and f0 maximum or a F1-F3 formant dispersion of "−1 Hz", etc. These were corrected my manual re-measurements (or deleted from the dataset).

2.6.3 Manual gesture measures

All manual communicative gestures were annotated by considering the gestural stroke (the most effortful part of the gesture, which usually constitutes its semantic unit; Kendon, 2004; McNeill, 1992; Rohrer et al., 2020). Non-communicative gestures such as self-adaptors (e.g., scratching, touching hair; Ekman and Friesen, 1969) were excluded. Gesture rate was calculated per speech as the total number of gestures produced relative to the phonation time in minutes (gestures/phonation time).
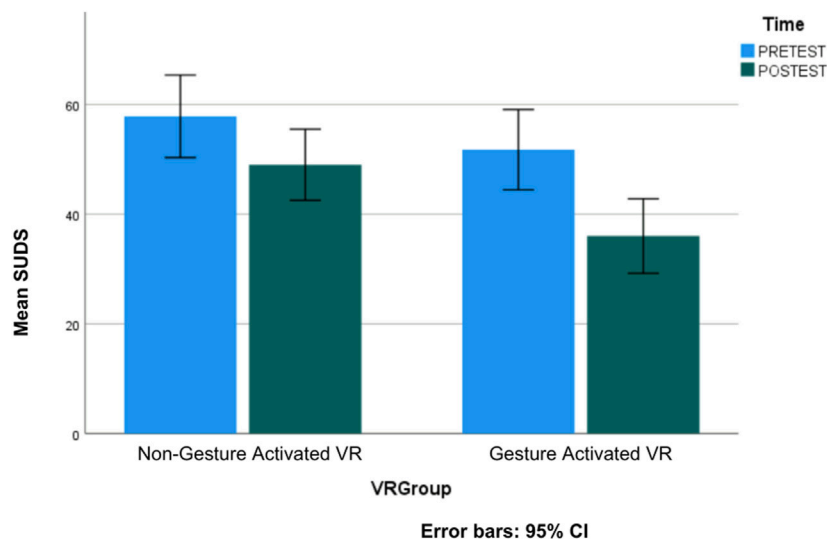
FIGURE 2
Mean SUDS values at pre- and post-training for both Non-Gesture Activated and Gesture Activated VR conditions.

2.7 Statistical analyses

Statistical analyses were performed using IBM SPSS Statistics 19. A number of GLMMs were run for the following independent variables, namely, self-perceived anxiety (SUDS), persuasiveness and charisma, and gesture rate, and a set of 20 values for all the prosodic parameters (5 for f0, 7 for duration and 8 for voice quality). All the GLMM models included Condition (two levels: Gesture Activated VR group and Non-Gesture Activated VR group) and Time (two levels: pre-training; 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.

2.8 Ethical approval

This study was approved by the Universitat Pompeu Fabra's Ethical Review Board for Research Projects (Comissió Institucional de Revisió Ètica de Projects CIREP-UPF) and also received approval from Recercaixa Project [2017 ACUP 00249]. Prior written informed consent was obtained from each participant and/or their parents or legal guardians, as appropriate.

3 Results

3.1 Self-assessed anxiety

The GLMM analysis for SUDS showed a main effect of Condition (F (1,140) = 4.805, p = .030), which indicated that in general (at both pre- and post-training) Non-Gesture Activated VR group values were higher than Gesture Activated VR group values (β = 10.071, SE = 4.595, p = .030), and a main effect of Time (F (1,140) = 41.889, p < .001), showing that SUDS values were lower at post-training regardless of the

condition (β = 12.381, SE = 1.913, p < .001). Also, a significant interaction between Condition and Time was obtained (F (1,140) = 4.474, p = .036). Post-hoc analyses revealed a significant difference between the two groups at post-training, showing a lower SUDS score for the Gesture Activated VR condition compared to the Non-Gesture Activated VR condition (β = 16.429, SE = 2.470, p < .001, g = 0.66). From pre- to post-training the Non-Gesture Activated VR condition significantly decreased their values: (β = 8.333, SE = 2.922, p = .005, g = 0.47), and so did the Gesture Activated VR condition: (β = 16.429, SE = 2.470, p < .001, g = 0.74). The graph in Figure 2 shows the mean SUDS scores separated by Condition (Gesture Activated VR group and Non-Gesture Activated VR group) and Time (pre-training and post-training). Table 1 displays the descriptive statistics for SUDS.

3.2 Perceived persuasiveness and charisma

The GLMM analysis for persuasiveness showed a near-significant main effect of Condition (F (1,112) = 3.778, p = .054, g =), which indicated that Non-Gesture Activated VR group values showed a tendency to be lower than Gesture Activated VR values (β = 7.281, SE = 3.746, p = .054), and a main effect of Time (F (1,112) = 24.552, p < .001), showing that persuasiveness values were higher at post-training independently of the condition (β = 4.588, SE = .909, p < .001). Also, a significant interaction between Condition and Time was obtained (F (1,112) = 4.560, p = .035). Post-hoc analyses revealed a significant difference between the two groups at post-training (β = 9.256, SE = 3.719, p = .014, g = 0.67), showing higher persuasiveness scores for the Gesture Activated VR condition. From pre- to post-training the Non-Gesture Activated VR condition significantly increased their values (β = 2.607, SE = 1.253, p = .04, g = 0.17), and so did the Gesture Activated VR condition: (β = 6.500, SE = 1.211, p < .001, g = 0.44). The graph in Figure 3 shows the mean persuasiveness scores separated by Condition (Gesture Activated VR group and Non-Gesture Activated VR group)

TABLE 1 Descriptive statistics for SUDS in each of the two conditions.

	Group	Session	M	SD	SE	95% CI
SUDS	Non-Gesture Activated VR	Pre-test	57.83	20.16	3.68	[50.31 65.36]
		Post-test	49	17.39	3.17	[42.51 55.49]
	Gesture Activated VR	Pre-test	51.75	22.88	3.61	[44.43 59.07]
		Post-test	36	21.24	3.36	[29.20 42.80]
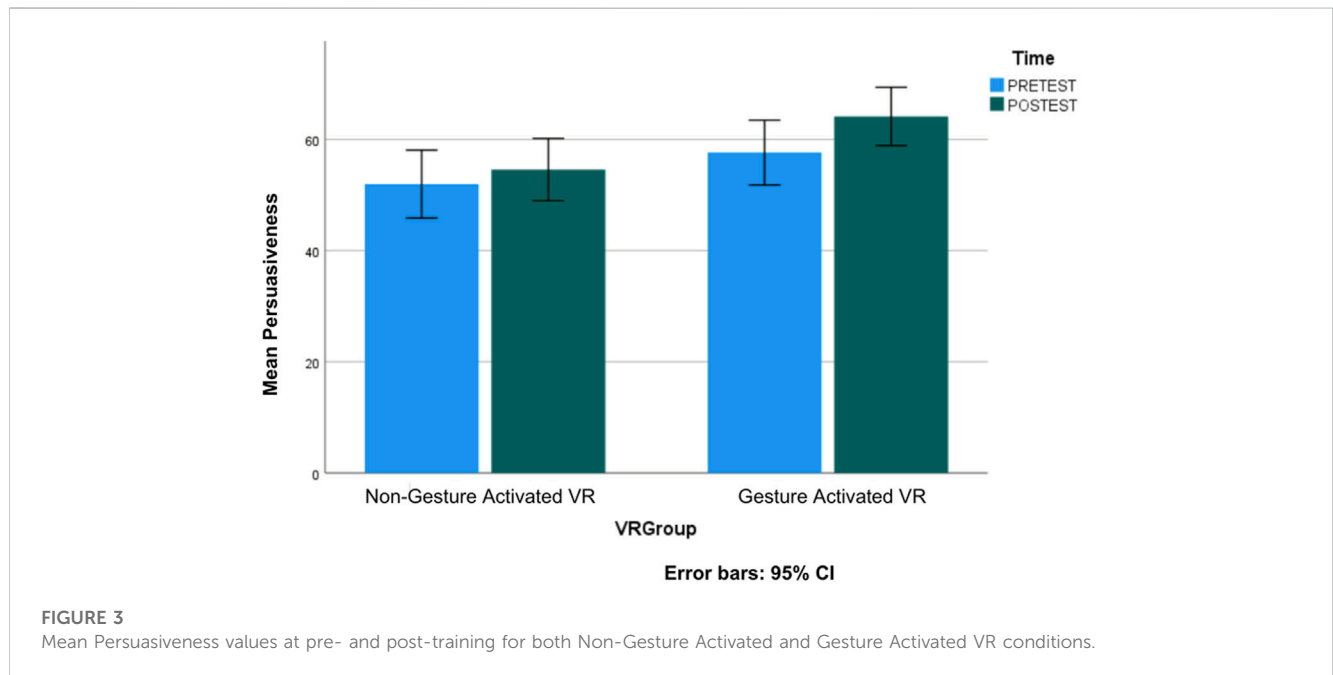
FIGURE 3
Mean Persuasiveness values at pre- and post-training for both Non-Gesture Activated and Gesture Activated VR conditions.

TABLE 2 Descriptive statistics for Persuasiveness in each of the two conditions.

	Group	Session	M	SD	SE	95% CI
Persuasiveness	Non-Gesture Activated VR	Pre-test	51.96	15.77	2.98	[45.85 58.08]
		Post-test	54.57	14.45	2.73	[48.97 60.17]
	Gesture Activated VR	Pre-test	57.63	15.58	2.84	[51.81 63.45]
		Post-test	64.13	14.07	2.56	[58.88 69.39]

and Time (pre-training and post-training). Table 2 displays the descriptive statistics for persuasiveness.

Regarding charisma, the GLMM analysis showed a main effect of Time (F (1,112) = 13.109, $p < .001$), which indicated that pre-training scores were lower for both conditions ($\beta = 2.945$, SE = .813, $p < .001$). The analysis also showed a significant interaction between Time and Condition (F (1,112) = 5.717, $p = .018$). Post-hoc analyses revealed a significant difference between the two groups at post-training ($\beta = 9.664$, SE = 3.813, $p = .013$, $g = 0.67$). From pre- to post-training the charisma scores of the Gesture Activated VR group were significantly higher than at pre-training: $\beta = 4.889$, SE = 1.139, $p < .001$, $g = 0.33$; by contrast, the charisma scores for the Non-Gesture Activated VR condition did not significantly differ from pre- to post-training. The

graph in Figure 4 shows the mean charisma scores separated by Condition (Gesture Activated VR group and Non-Gesture Activated VR group) and Time (pre-training and post-training). Table 3 displays the descriptive statistics for charisma.

3.3 Prosodic parameters

3.3.1 F0

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 4 shows the results of those GLMM analyses in terms of main effects
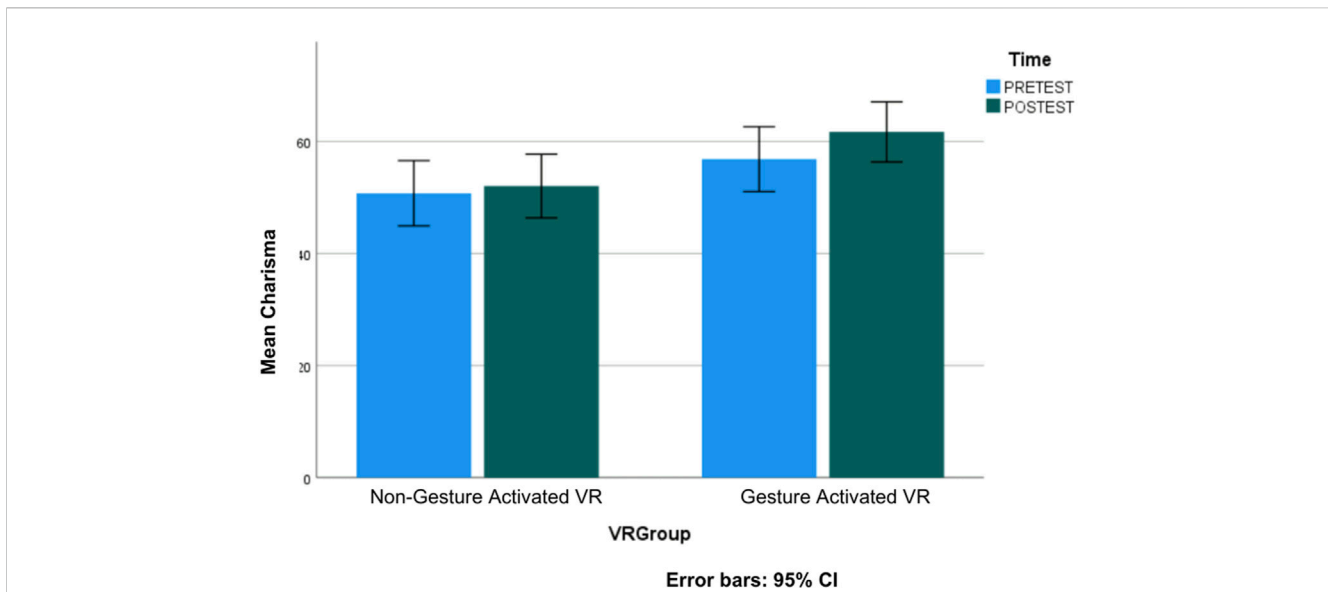
FIGURE 4
Mean Charisma values at pre- and post-training for both Non-Gesture Activated and Gesture Activated VR conditions.

TABLE 3 Descriptive statistics for Charisma in each of the two conditions.

	Group	Session	M	SD	SE	95% CI
Charisma	Non- Gesture Activated VR	Pre-test	50.75	15.02	2.84	[44.93 56.57]
		Post-test	52.04	14.65	2.76	[46.35 57.72]
	Gesture Activated VR	Pre-test	56.83	15.49	2.83	[51.05 62.62]
		Post-test	61.7	14.37	2.62	[56.33 67.07]

(Time and Condition), as well as interactions between Time and Condition. Summarizing, a main effect of Time was obtained only for f0 maximum, meaning that the post-training values in both groups were lower than the pre-training values. A main effect of Condition was only obtained for f0 mean, meaning that the participants in the Gesture Activated VR group produced lower f0 values across both pre- and post-training phases. No significant interactions were obtained for any of the variables.

3.3.2 Tempo

Regarding tempo, a set of seven 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 5 shows the results 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 number of syllables. A main effect of Condition was obtained for four variables, namely, number of silent pauses, speech rate, net syllable rate and ASD, meaning that the participants in the Gesture Activated VR group had lower speech-rate and net-syllable-rate (or articulation-rate) values, as well as higher ASD values. No significant interactions emerged for this domain either.

3.3.3 Voice quality

In the domain of voice quality measurements, a set of eight GLMMs were applied to our target variables, namely, h1*-h2*, h1-A3, CPP, Harmonicity, CoG, formant dispersion 1–3, shimmer, and jitter. Table 6 shows the results 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 for six variables, namely, h1-A3, CPP, CoG, formant dispersion 1-3, shimmer, and harmonicity, meaning that pre-training values were lower across groups for all the variables except for CoG and shimmer. A main effect of Condition was obtained for four variables, namely, h1*-h2*, h1-A3, shimmer, and jitter, meaning that the participants in the Gesture Activated VR group produced lower values compared to the Non-Gesture Activated VR group, both at pre- and post-training. No significant interactions were found for any of the variables.

3.4 Manual gesture rate

To assess whether the additional embodiment instruction given to the participants of the Gesture Activated VR group was effective, we counted the number of manual gestures

TABLE 4 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	$F(1,113) = .036, p = .850$	$F(1,113) = 5.710, p = .019$	$F(1,113)= .497, p = .482$
f0 max	$F(1,114) =4.562, p = .035$	$F(1,114) = 6.117, p = .015$	$F(1,114)= 1.717, p = .193$
f0 variability	$F(1,114) = .308, p = .580$	$F(1,114) = .533, p = .467$	$F(1,114) = 3.253, p = .074$
f0 mean	$F(1,116) = .039, p = .844$	$F(1,116) = 8.414, p = .004$	$F(1,122)= 1.022, p = .314$
f0 range	$F(1,114) =2.202, p =.141$	$F(1,114) = .349, p = .556$	$F(1,114)= .186, p = .667$

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

Variable	Main effect of Time	Main effect of Condition	Interaction Time*Condition
Number of syllables	$F(1,114) = 7.150, p = .009$	$F(1,114) = 2.969, p= .088$	$F(1,114) = 2.074, p = .153$
Number of silent pauses	$F(1,114) = .059, p = .809$	$F(1,114) = 11.119, p = .001$	$F(1,114) = .567, p=.453$
Total time of the presentation	$F(1,116) = 3.535, p = .063$	$F(1,116) = .229, p = .696$	$F(1,116) = .020, p = .889$
Total speaking time	$F(1,116) = 1.511, p = .221$	$F(1,116) = 3.661, p = .058$	$F(1,116) = 1.881, p = .173$
Speech rate	$F(1,116) = 1.306, p = .256$	$F(1,116) = 4.401, p = .038$	$F(1,116) = 2.215, p = .139$
Net syllable rate	$F(1,114) = .090, p = .765$	$F(1,114) = 6.378, p = .013$	$F(1,114) = .832, p=.363$
ASD	$F(1,112) = .712, p = .401$	$F(1,112) = 27.377, p < .001$	$F(1,112) = 1.375, p= .244$

TABLE 6 Summary of the GLMM analyses for the 8 voice variables, in terms of main effects and interactions.

Variable	Main effect of Time	Main effect of Condition	Interaction Time*Condition
h1*–h2*	$F(1,110) = .195, p = .659$	$F(1,110) = 8,478, p = .004$	$F(1,110) = .633, p = .428$
h1-A3	$F(1,110) = 10.927, p = .001$	$F(1,110) = 8.247, p = .005$	$F(1,110) = .730, p = .395$
CPP	$F(1,110) = 13.428, p < .001$	$F(1,110) = .000, p = .997$	$F(1,110) = .382, p = .538$
Harmonicity	$F(1,110) = 9.216, p = .003$	$F(1,110) = .061, p = .806$	$F(1,110) = 1.671, p = .199$
CoG	$F(1,110) = 31.521, p < .001$	$F(1,110) = 2.653, p = .106$	$F(1,110) = .220, p = .640$
Formant dispersion 1–3	$F(1,110) = 5.813, p = .018$	$F(1,110) = .005, p = .945$	$F(1,110) = .975, p = .326$
Shimmer	$F(1,110) = 4.248, p = .042$	$F(1,110) = 30.494, p < .001$	$F(1,110) = .194, p = .660$
Jitter	$F(1,110) = 2.926, p = .090$	$F(1,110) = 22.931, p < .001$	$F(1,110) = .422, p = .517$

performed by participants in both conditions during their pre-training speech, as well as in the first and third VR-assisted training sessions. As noted above, gesture rate was calculated as the total number of hand gestures produced relative to the phonation time in minutes. The results showed that the mean gesture rate at pre-training was 42.27 gestures per minute for the Non-Gesture Activated VR group and 32.89 gestures per minute for the Gesture Activated VR group. For training sessions 1 and 3, the mean gesture rates were 28.53 gestures per minute for the Non-Gesture Activated VR group and 25.27 per minute for the Gesture Activated VR group. Crucially, the difference from pre-training to training session 1 was a reduction of 13.74 for the Non-Gesture Activated VR group compared with a reduction of only 7.62 for the Gesture Activated VR group. These results clearly indicate that Gesture Activated VR participants maintained their gesture rate when they underwent the training sessions, their relative use of

manual gestures being higher than that of the Non-Gesture Activated VR participants.

A GLMM was applied to this data. A main effect of Time was obtained ($F(1,114) = 4.276, p = .041$), meaning that at post-training values were higher across groups ($\beta = 2.895$, SE $= 1.400, p = .041$). A main effect of Condition was also obtained ($F(1,114) = 10.144, p = .002$), meaning that Gesture Activated VR scores were higher across both pre- and post-training phases ($\beta = 11.229$, SE $= 3.167, p = .001$).

4 Discussion and conclusion

The central aim of the study was to investigate whether explicitly instructing secondary students to use gesture during a three-session VR-assisted public speaking training program would help reduce their levels of PSA and, in addition, enhance the quality of their

performance in front of a small live audience after training. Therefore, a between-subjects experiment with a pre-training speech, three training sessions, and a post-training speech was designed so that we could compare pre-to post-training speeches between a group of students instructed to embody their speeches while speaking to the VR audience and a group who received no such instruction. One of the key features of the study was that it included a comprehensive assessment of the students' public speaking performance before and after their VR-assisted training sessions. Specifically, the study assessed whether presenters giving their post-training speech reported lower levels of anxiety and displayed higher levels of persuasiveness and charisma, and/or produced a more audience-oriented speech from the point of view of prosodic and gestural features. In order to make the VR technology accessible to everyone, the study utilized a cost-effective method consisting of cardboard glasses attached to a phone that allowed us to recommend the application to students and instructors who showed interest in practicing their public speaking after the completion of the experiment at home and at school when needed.

In relation to the effects on anxiety, our results showed a significant reduction in the degree of anxiety in both Non-Gesture Activated VR and Gesture Activated VR conditions. Firstly, these results support previous VR training studies that reported a reduction in the self-assessed PSA levels of participants in clinical (e.g., Lister, 2016; Lindner et al., 2018; Yuen et al., 2019; Zacarin et al., 2019) and educational settings (e.g., Heuett and Heuett, 2011; Kahlon et al., 2019). Second, a key finding of the study is that the embodiment prompt during the VR training sessions triggered a significantly stronger effect in the reduction of self-perceived anxiety among participants in this condition as compared with the participants in the Non-Gesture Activated VR condition. These results expand previous findings on the positive effects that physical activity has on mental health (i.e., wellbeing and self-concept, as reported in McDonald and Hodgdon, 1991; Fox, 2000) and cognitive functioning (see Donnelly et al., 2016 for a review), as well as on the reduction of anxiety (e.g., Korczak et al., 2017).

Focusing now on the effects of embodiment on persuasiveness and charisma, a key finding of the present study is that the participants in the Gesture Activated VR condition increased their persuasiveness and charisma ratings from pre-to post-training, as opposed to the participants in the Non-Gesture Activated VR condition. Perceptual ratings of persuasiveness and charisma were used, as has been done in previous studies analyzing speakers' persuasiveness or charisma (e.g., Maricchiolo et al., 2010; Jackob et al., 2011; Berger et al., 2017), a very high level of inter-rater reliability having been confirmed.

The present results seem to be connected to recent findings from research showing that the activation of the body and gesturing while performing speaking tasks has direct consequences on speakers' cognitive processes because it helps speakers to reduce the amount of cognitive resources they need to formulate speech (Wagner et al., 2004), enhances their problem-solving abilities (Thomas and Lleras, 2009), and improves their ability to retain memories of things they have just learned (e.g., Alibali and Goldin-Meadow, 1993). Along these lines, we contend that our results constitute further evidence in support of the embodied cognition paradigm as a successful way to encourage learning through the activation of the body. As studies from numerous fields in neuroscience, linguistics, and cognitive

science have claimed, "the highest percentage of human cognitive ability is based on bodily capabilities to produce knowledge" (Kosmas, 2019: 3) (see also Wilson, 2002; Gallese and Lakoff, 2005). We can speculate that by reminding participants to use their bodies to enhance their expressiveness, the speeches produced by the Gesture Activated VR group may have been enriched by this awareness of the body as a tool for the construction of effective discourse (Kalantzis and Cope, 2004). Moreover, this body activation may have favored a stronger feeling of self-confidence that was key to rater perceptions that they were more charismatic speakers and their messages more persuasive (McDonald and Hodgdon, 1991; Fox, 2000).

Another important factor that might explain the positive results obtained by Gesture Activated VR participants is the relationship between body movement and the greater sense of presence they perhaps experienced in the simulated VR environment. Following up on previous results (e.g., Slater et al., 1995; Slater et al., 1998), the fact that participants in the Gesture Activated VR condition received the instruction to use their body to increase their expressiveness could have enhanced their sense of presence and the VR experience could have been more immersive to them than to participants in the other condition (Bianchi-Berthouze, 2013). Encouraging participants to use their bodies could have triggered a more realistic and vivid VR experience, and this sense of enhanced presence was then transferred to the post-training live audience context, since crucially speakers in this group were perceived as more persuasive and charismatic. Although the study did not include any measure of presence, in our view it would be interesting to include this measure in future studies in order to analyze its relationship with gesture use and embodiment measures.

Regarding the effects of the Gesture Activated VR condition on prosodic parameters, significant interactions were obtained neither for f0 and tempo nor voice quality parameters, meaning that the addition of an embodiment instruction while employing VR did not lead to any differences in these prosodic parameters in the pre- and post-training speeches. These results contradict our expectations, given the reported relation between the prosodic features of speeches and their persuasiveness (e.g., Kelly and Goldsmith, 2004; Maricchiolo et al., 2009; Jackob et al., 2011; Yokoyama and Daibo, 2012; Peters and Hoetjes, 2017). Nevertheless, a possible explanation for the lack of significant changes in the prosodic parameters in post-training speeches is that already in the pre-training session the Gesture Activated VR group showed significant differences in the majority of the prosodic parameters compared to the Non-Gesture Activated VR group. These differences suggest that the Gesture Activated VR group had a higher level of audience-orientation right from the start and kept that high level also after training. That is, the Gesture Activated VR group was already performing well while the Non-Gesture Activated VR group was not able to improve further, which, in combination, prevented interaction effects from emerging. With regard to the five f0 values, no significant melodic changes were observed between the pre- and post-training speeches across groups. Even though no significant interactions were found, the Gesture Activated VR group showed a general tendency to produce a less thin and breathy but more harmonious and sonorous voice, key attributes of speech perceived as charismatic. The Gesture Activated VR group also used fewer pauses and a reduced net syllable rate, which is consistent

with the listener-oriented speaking style that makes speeches more persuasive and is likely to signal greater credibility (Jackob et al., 2011).

Focusing on the gesture rate used in the pre-to post-training speeches, no significant differences were found across conditions. We expected to observe a significantly higher rate of gesturing in the Gesture Activated VR group because of the explicit instruction they had received in that regard. Though there was a higher relative increase in gesture rate at post-training for the Gesture Activated VR group, the difference between the two conditions was not significant. Therefore, our hypothesis regarding a higher gesture rate for the Gesture Activated VR condition was not supported. Interestingly, however, the fact that the embodiment instruction did not cause participants to perform significantly more gestures at post-training is consistent with the results of previous studies showing that the most effective and credible speaking style is characterized not by a very extensive use of gesture but rather by a moderate one (e.g., Dargue et al., 2019; Rodero, 2022; Rodero et al., 2022).

In summary, the results of our prosodic and gesture analyses of the student-produced speeches revealed no significant differences in prosodic or gesture parameters across groups. This is somewhat surprising given the fact that significant gains were obtained in perceived persuasiveness and charisma in the embodied condition. We expected to see some correlations between a more charismatic style and an increase in discourse persuasiveness in terms of the use of specific prosodic and gestural parameters. Gesture rate, then, might not be a suitable measure of a speaker's overall multimodal behavior, which involves also gesture amplitude and timing in co-creating communicative meanings together with prosody as well as a bundle of features such as eye gaze patterns, facial expressions, and body posture (Signorello et al., 2012). This suggests that further and more detailed analyses of multimodal behavior would be needed for this data.

In summary, we can conclude that explicitly instructing students to use gestures when they are practicing public speaking in a VR-assisted environment has the potential to boost some of the performance parameters after intervention, when the students are asked to speak before a live audience. Specifically, it can help make the students less anxious, as well as more charismatic and persuasive. Our results have important educational implications. First, they confirm the value of applying VR technology in the classroom to enable students to practice developing their oral skills, in the process increasing their self-confidence and awareness of their oral communicative strengths (e.g., Van Ginkel et al., 2019), thereby leading to more charismatic delivery (Niebuhr and Michalsky, 2018; Niebuhr and Tegtmeier, 2019). Second, they show that adding embodiment instructions as a complementary technique can augment the positive effects of VR-assisted training on subsequent public speaking tasks. In general, our results confirm and expand previous results on the positive value of embodied learning approaches in language education: not only can embodied learning add emotional and motivational value benefits to language learning contexts by virtue of the fact that physical activities make classroom learning more enjoyable (Hanks and Eckstein, 2019; Kosmas and Zaphiris, 2019; see Jusslin et al., 2022 for a review). It also heightens student interest, overall wellbeing, and self-confidence (Mathias and von Kriegstein, 2022; Cannon, 2017; Hanks and Eckstein, 2019).

Several limitations must be considered. First of all, the study was conducted with a sample of 17-year-old students and the results

cannot be safely generalized to other age groups, as PSA could vary with age. Nor were our two groups of participants controlled for in terms of gender, and it would have been interesting to assess possible differences between genders in the outcomes obtained.

Second, participants could not see their hands—either real or virtual—as they performed their speeches, which may have inhibited or otherwise distorted their embodiment behavior. Being able to see virtual hands and/or full virtual body would contribute to the sense of presence experienced by participants, something that we did not measure here. The sense of virtual ownership that users can experience seeing their virtual bodies in the VR environments could not take place in this study, as the VR application utilized did not feature it. Future research could implement the gesture-encouraging condition with a VR application that includes this feature.

Third, though anxiety levels were measured, the instrument used depended on self-reporting. Although SUDS has been widely used in public speaking studies and represents a validated overall measure of emotional distress (e.g., Tanner, 2012), adding objective measures such as electrophysiological data would allow us to obtain a more fine-grained picture of participant anxiety levels and compare them with other measures. Also, our analyses of persuasiveness and charisma would have been more comprehensive had they included an assessment of the cogency of the arguments deployed by speakers. And as we have noted, considerable work needs to be done to clarify the relationship between persuasiveness and charisma on the one hand and prosodic and gestural features on the other.

Finally, future longitudinal studies could be carried out in which public speaking practice before VR-simulated audiences takes place over more or longer sessions, possibly in combination with various feedback strategies.

In conclusion, the results of the present investigation offer further hints on how VR-simulated environments can be most effectively used by secondary students to sharpen their public speaking skills. Specifically, they show that the addition of a brief embodiment instruction suggesting that speakers combine their oral performance with the use of gestures not only seems to make for a more vivid VR experience but possibly also leads to reduced anxiety and concomitant gains in public speaking performance. These results have important academic implications, suggesting as they do that VR technology can be profitably employed as a complementary and powerfully engaging tool for the teaching of oral communication at the secondary school level.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Universitat Pompeu Fabra's Ethical Review Board for Research Projects (Comissió Institucional de Revisió Ètica de Projects CIREP-UPF). The studies were conducted in accordance with the local legislation

and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

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

Conflict of interest

The 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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frvir.2023.1074062/full#supplementary-material

References

Addington, D. W. (1971). The effect of vocal variations on ratings of source credibility. *Speech Monogr.* 38, 242–247. doi:10.1080/03637757109375716

Adler, R. B. (1980). Integrating reticence management into the basic communication curriculum. *Commun. Educ.* 29, 215–221. doi:10.1080/03634528009378415

Alibali, M. W. (2005). Gesture in spatial cognition: expressing, communicating, and thinking about spatial information. *Spat. Cogn. Comput.* 5, 307–331. doi:10.1207/s15427633scc0504_2

Alibali, M. W., and Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: what the hands reveal about a child's state of mind. *Cogn. Psychol.* 25 (4), 468–523. doi:10.1006/cogp.1993.1012

Anderson, C., and Galinsky, A. (2006). Power, optimism, and risk-taking. *Eur. J. Soc. Psychol.* 36, 511–536. doi:10.1002/ejsp.324

Armel, K. C., and Ramachandran, V. S. (2003). Projecting sensations to external objects: evidence from skin conductance response. *Proc. R. Soc. Lond. Biol. Sci.* 270, 1499–1506. doi:10.1098/rspb.2003.2364

Ayres, J., and Hopf, T. S. (1985). Visualization: A means of reducing speech anxiety. *Commun. Educ.* 34 (4), 318–323. doi:10.1080/03634528509378623

Bäckström, T., Lehto, L., Alku, P., and Vilkman, E. (2003). Automatic pre-segmentation of running speech improves the robustness of several acoustic voice measures. *Logop. Phoniatr. vocology* 28 (3), 101–108. doi:10.1080/14015430310015237

Bailey, E. (2018). A historical view of the pedagogy of public speaking. *Voice Speech Rev.* 13 (1), 31–42. doi:10.1080/23268263.2018.1537218

Banse, R., and Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *J. personality Soc. Psychol.* 70 (3), 614–636. doi:10.1037/0022-3514.70.3.614

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660. doi:10.1017/s0140525x99002149

Bartholomay, E. M., and Houlihan, D. D. (2016). Public speaking anxiety scale: preliminary psychometric data and scale validation. *Personality Individ. Differ.* 94, 211–215. doi:10.1016/j.paid.2016.01.026

Berger, S., Niebuhr, O., and Peters, B. (2017). "Winning over an audience – a perception-based analysis of prosodic features of charismatic speech," in Proceedings of the 43rd Annual Conference of The German Acoustical Society, Kiel, Germany, April 2017, 1454–1457.

Bianchi-Berthouze, N., Kim, W. W., and Patel, D. (2007). "Does body movement engage you more in digital game play? And why?," in *Affective computing and intelligent interaction* (Berlin, Heidelberg: Springer). doi:10.1007/978-3-540-74889-2_10

Bianchi-Berthouze, N. (2013). Understanding the role of body movement in player engagement. *Human–Computer Interact.* 28 (1), 40–75. doi:10.1080/07370024.2012.688468

Blume, B. D., Dreher, G. F., and Baldwin, T. T. (2010). Examining the effects of communication apprehension within assessment centres. *J. Occup. Organ. Psychol.* 83 (3), 663–671. doi:10.1348/096317909x463652

Boersma, P., and Weenink, D. (2007). *Praat: Doing phonetics by computer*.

Boetje, J., and van Ginkel, S. (2020). The added benefit of an extra practice session in virtual reality on the development of presentation skills: A randomized control trial. *J. Comput. Assisted Learn.* 37 (1), 253–264. doi:10.1111/jcal.12484

Bowman, D., and Hodges, L. (1999). Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments. *J. Vis. Lang. Comput.* 10 (1), 37–53. doi:10.1006/jvlc.1998.0111

Boyce, J. S., Alber-Morgan, S. R., and Riley, J. G. (2007). Fearless public speaking. *Child. Educ.* 83 (3), 142–150. doi:10.1080/00094056.2007.10522899

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:10.1016/j.landurbplan.2015.02.005

Burgmer, P., and Englich, B. (2013). Bullseye!: how power improves motor performance. *Soc. Psychol. Personality Sci.* 4 (2), 224–232. doi:10.1177/1948550612452014

Burgoon, J. K., Birk, T., and Pfau, M. (1990). Nonverbal behaviors, persuasion, and credibility. *Hum. Commun. Res.* 17 (1), 140–169. doi:10.1111/j.1468-2958.1990.tb00229.x

Cannon, A. (2017). When statues come alive: teaching and learning academic vocabulary through drama in schools. *TESOL Q.* 51 (2), 383–407. doi:10.1002/tesq.344

Carney, D., Cuddy, A., and Yap, A. (2010). Power posing: brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychol. Sci.* 21, 1363–1368. doi:10.1177/0956797610383437

Church, R. B., and Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition* 23 (1), 43–71. doi:10.1016/0010-0277(86)90053-3

Cook, S. W., and Goldin-Meadow, S. (2006). The role of gesture in learning: do children use their hands to change their minds? *J. Cognition Dev.* 7 (2), 211–232. doi:10.1207/s15327647jcd0702_4

Cuddy, A., Wilmuth, C. A., and Carney, D. R. (2012). *The benefit of power posing before a high-stakes social evaluation*. Harvard Business School Working Paper, No 13-027.

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

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 2019 (ICITDA 2019), Yogyakarta, Indonesia, November 2019.

Dargue, N., Sweller, N., and Jones, M. P. (2019). When our hands help us understand: A meta-analysis into the effects of gesture on comprehension. *Psychol. Bull.* 145 (8), 765–784. doi:10.1037/bul0000202

Darwin, C. R. (1872). *The expression of the emotions in man and animals*. London: John Murray.

Davis, M., Papini, S., Rosenfield, D., Roelofs, K., Kolb, S., Powers, M., et al. (2017). A randomized controlled study of power posing before public speaking exposure for social anxiety disorder: no evidence for augmentative effects. *J. Anxiety Disord.* 52, 1–7. doi:10.1016/j.janxdis.2017.09.004

De Jong, N. H., and Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behav. Res. methods* 41 (2), 385–390. doi:10.3758/brm.41.2.385

Donnelly, J. E., Hillman, C. H., Castelli, D., Etnier, J. L., Lee, S., Tomporowski, P., et al. (2016). Physical activity, fitness, cognitive function, and academic achievement in children: A systematic review. *Med. Sci. sports Exerc.* 48 (6), 1197–1222. doi:10.1249/MSS.0000000000000901

Ekman, P., Friesen, W. V., and Scherer, K. R. (1976). Body movement and voice pitch in deceptive interaction. *Semiotica* 26, 23–27. doi:10.1515/semi.1976.16.1.23

Ekman, P., and Friesen, W. V. (1969). The repertoire of nonverbal behavior: categories, origins, usage, and coding. *Semiotica* 1, 49–98. doi:10.1515/semi.1969.1.1.49

Feyereisen, P., and Havard, I. (1999). Mental imagery and production of hand gestures while speaking in younger and older adults. *J. Nonverb. Behav.* 23, 153–171. doi:10.1023/A:1021487510204

Fox, K. R. (2000). Self-esteem, self-perceptions and exercise. *Int. J. Sport Psychol.* 31, 228–240.

Gallese, V., and Lakoff, G. (2005). The brain's concepts: the role of the sensory-motor system in conceptual knowledge. *Cogn. Neuropsychol.* 22, 455–479. doi:10.1080/02643290442000310

Gao, D. (2022). Oral English training based on virtual reality technology. *Eng. Intell. Syst.* 30 (1), 49–54.

Gnisci, A., and Pace, A. (2014). The effects of hand gestures on psychosocial perception: A preliminary study. *Smart Innovation, Syst. Technol.* 26, 305–314. doi:10.1007/978-3-319-04129-2_30

Gunnell, K. E., Flament, M. F., Buchholz, A., Henderson, K. A., Obeid, N., Goldfield, G. S., et al. (2016). Examining the bidirectional relationship between physical activity, screen time, and symptoms of anxiety and depression over time during adolescence. *Prev. Med.* 88, 147–152. doi:10.1016/j.ypmed.2016.04.002

Hall, J. A., Coats, E. J., and Smith LeBeau, L. (2005). Nonverbal behavior and the vertical dimension of social relations: A meta-analysis. *Psychol. Bull.* 131, 898–924. doi:10.1037/0033-2909.131.6.898

Hanks, E., and Eckstein, G. (2019). Increasing English learners' positive emotional response to learning through dance. *TESL Report.* 52 (1), 72–93.

Heuett, B. L., and Heuett, K. B. (2011). Virtual reality therapy: A means of reducing public speaking anxiety. *Int. J. Humanit. Soc. Sci.* 1, 1–6. doi:10.3390/jpm10010014

Hostetter, A. B., and Alibali, M. W. (2019). Gesture as simulated action: revisiting the framework. *Psychonomic Bull. Rev.* 26 (3), 721–752. doi:10.3758/s13423-018-1548-0

Hostetter, A. B., and Alibali, M. W. (2004). "On the tip of the mind: gesture as a key to conceptualization," in *Proceedings of the 26th annual conference of the cognitive science society*. Editors K. Forbus, D. Gentner, and T. Regier (Mahwah, NJ: Erlbaum), 589–594.

Hostetter, A. B., and Alibali, M. W. (2008). Visible embodiment: gestures as simulated action. *Psychonomic Bull. Rev.* 15, 495–514. doi:10.3758/PBR.15.3.495

Jackob, N., Roessing, T., and Petersen, T. (2011). The effects of verbal and nonverbal elements in persuasive communication: findings from two multi-method experiments. *Communications* 36 (2), 245–271. doi:10.1515/COMM.2011.012

Jusslin, S., Korpinen, K., Lilja, N., Martin, R., Lehtinen-Schnabel, J., and Anttila, E. (2022). Embodied learning and teaching approaches in language education: A mixed studies review. *Educ. Res. Rev.* 37, 100480. doi:10.1016/j.edurev.2022.100480

Kougl, K. M. (1980). Dealing with quiet students in the basic college speech course. *Commun. Educ.* 29, 234–238. doi:10.1080/03634528009378418

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 (1), 47. doi:10.1186/s13034-019-0307-y

Kalantzis, M., and Cope, B. (2004). Designs for learning. *E-Learning Digital Media* 1 (1), 38–93. doi:10.2304/elea.2004.1.1.7

Kelly, S. D., and Goldsmith, L. H. (2004). Gesture and right hemisphere involvement in evaluating lecture material. *Gesture* 4 (1), 25–42. doi:10.1075/gest.4.1.03kel

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kilteni, K., Groten, R., and Slater, M. (2012). The sense of embodiment in virtual reality. *Presence Teleoperators Virtual Environ.* 21, 373–387. doi:10.1162/PRES_a_00124

Kita, S. (2000). "How representational gestures help speaking," in *Language and gesture*. Editor D. McNeill (Cambridge: Cambridge University Press), 162–185.

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

Korczak, D. J., Madigan, S., and Colasanto, M. (2017). Children's physical activity and depression: A meta-analysis. *Pediatrics* 139 (4), e20162266. doi:10.1542/peds.2016-2266

Kosmas, P., and Zaphiris, P. (2019). Words in action: investigating students' language acquisition and emotional performance through embodied learning. *Innovation Lang. Learn. Teach.* 14 (4), 317–332. doi:10.1080/17501229.2019.1607355

Krauss, R. M., Chen, Y., and Chawla, P. (1996). "Nonverbal behavior and nonverbal communication: what do conversational hand gestures tell us?," in *Advances in experimental social psychology*. Editor M. P. Zanna (Cambridge, Massachusetts: Academic Press), 389–450. doi:10.1016/S0065-2601(08)60241-5

Krauss, R. M., Chen, Y., and Gottesman, R. F. (2000). "Lexical gestures and lexical access: A process model," in *Language and gesture*. Editor D. McNeill (Cambridge: Cambridge University Press), 261–283. doi:10.1017/cbo9780511620850.017

Kryston, K., Goble, H., and Eden, A. (2021). Incorporating virtual reality training in an introductory public speaking course. *J. Commun. Pedagogy* 4, 131–151. doi:10.31446/jcp.2021.1.13

Latu, I. M., Duffy, S., Pardal, V., and Alger, M. (2017). Power vs. persuasion: can open body postures embody openness to persuasion? *Compr. Results Soc. Psychol.* 2 (1), 68–80. doi:10.1080/23743603.2017.1327178

LeFebvre, L. E., LeFebvre, L., and Allen, M. (2020). Imagine all the people: imagined interactions in virtual reality when public speaking. *Imagination, Cognition Personality* 40, 189–222. doi:10.1177/0276236620938310

Legault, J., Zhao, J., Chi, Y., Chen, W., 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., Miloff, A., Fagernäs, 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. A., Piercey, 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. Rehabilitation* 3, 375–381.

Lister, H. A. (2016). *The effect of virtual reality exposure on fear of public speaking using cloud-based software*. Thesis. Heather Lister: University of New Brunswick.

Liu, X., and Xu, Y. (2014). "Body size projection and its relation to emotional speech—evidence from Mandarin Chinese," in Proceedings of Speech Prosody, Dublin, January 2014, 974–977.

Lövgren, T., and Doorn, J. V. (2005). "Influence of manipulation of short silent pause duration on speech fluency," in Proceedings of the Disfluency in Spontaneous Speech Conference, Aix-en-Provence, France, February 2005, 123–126.

Manusov, V., and Patterson, M. L. (2006). *The SAGE handbook of nonverbal communication*. Thousand Oaks, CA, USA: SAGE Publications. doi:10.4135/9781412976152

Maricchiolo, F., Gnisci, A., Bonaiuto, M., and Ficca, G. (2009). Effects of different types of hand gestures in persuasive speech on receivers' evaluations. *Lang. cognitive Process.* 24, 239–266. doi:10.1080/01690960802159929

Mathias, B., and von Kriegstein, K. (2023). Enriched learning: behavior, brain, and computation. *Trends Cognitive Sci.* 27 (1), 81–97. doi:10.1016/j.tics.2022.10.007

McDonald, D. G., and Hodgdon, J. A. (1991). *The psychological effects of aerobic fitness training: Research and theory*. NewYork: Springer-Verlag.

McMahon, E. M., Corcoran, P., O'Regan, G., Keeley, H., Cannon, M., Carli, V., et al. (2017). Physical activity in European adolescents and associations with anxiety, depression and well-being. *Eur. Child Adolesc. Psychiatry* 26 (1), 111–122. doi:10.1007/s00787-016-0875-9

McNeill, D. (2005). *Gesture and thought*. University of Chicago Press. Chicago, IL, USA, doi:10.7208/chicago/9780226514642.001.0001

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

Mehrabian, A., and Williams, M. (1969). Nonverbal concomitants of perceived and intended persuasiveness. *J. Personality Soc. Psychol.* 13, 37–58. doi:10.1037/h0027993

Michalsky, J., and Niebuhr, O. (2019). Myth busted? challenging what we think we know about charismatic speech. *AUC Philologica* 2019 (2), 27–56.

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. Adm.* 29 (1), 125. doi:10.1080/03634520701861713

Niebuhr, O., and Michalsky, J. (2018). "*Virtual reality simulations as a new tool for practicing presentations and refining public-speaking skills*".

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. Speech and computer*. Editors A. Karpov and R. Potapova (New York: Springer Nature), 357–368.

Niebuhr, O., and Tegtmeier, S. (2019). "Virtual reality as a digital learning tool in entrepreneurship: how virtual environments help entrepreneurs give more charismatic investor pitches," in *FGF studies in small business and entrepreneurship. Digital entrepreneurship* (Berlin, Germany: Springer), 123–158.

North, M. M., North, S. M., and Coble, J. R. (1998). Virtual reality therapy: an effective treatment for the fear of public speaking. *Int. J. Virtual Real.* 3 (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. *Proc. IS&T Int'l. Symp. Electron. Imaging Image Qual. Syst. Perform. XVIII* 2021, 294–294. doi:10.2352/ISSN.2470-1173.2021.9.IQSP-294

Pallavicini, F., and Pepe, A. (2020). Virtual reality games and the role of body involvement in enhancing positive emotions and decreasing anxiety: within-subjects pilot study. *JMIR serious games* 8 (2), e15635. doi:10.2196/15635

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

Peters, J., and Hoetjes, M. (2017). "The effect of gesture on persuasive speech," in Proceedings of the Interspeech 2017, 8th Annual Conference of the International Speech Communication Association, Stockholm, Sweden, August 2017, 659–663. doi:10.21437/Interspeech.2017-194

Pine, K. J., Lufkin, N., and Messer, D. (2004). More gestures than answers: children learning about balance. *Dev. Psychol.* 40 (6), 1059–1067. doi:10.1037/0012-1649.40.6.1059

Ranehill, E., Dreber, A., Johannesson, M., Leiberg, S., Sul, S., and Weber, R. A. (2015). Assessing the robustness of power posing: no effect on hormones and risk tolerance in a large sample of men and women. *Psychol. Sci.* 26 (5), 653–656. doi:10.1177/0956797614553946

Rasipuram, S., Rao, S. P., and Jayagopi, D. B. (2016). "Automatic prediction of fluency in interface-based interviews," in Proceedings of the 2016 IEEE Annual India Conference (INDICON), Bangalore, India, December 2016 (IEEE), 1–6.

Ray, G. (1986). Vocally cued personality prototypes: an implicit personality theory approach. *Commun. Monogr.* 53, 266–276. doi:10.1080/03637758609376141

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 (4), 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 (5), 749–760. doi:10.1177/0956797617744797

Rodero, E. (2022). Effectiveness, attractiveness, and emotional response to voice pitch and hand gestures in public speaking. *Front. Commun.* 7, 869084. doi:10.3389/fcomm.2022.869084

Rodero, E., Larrea, O., Rodríguez-de-Dios, I., and Lucas, I. (2022). The expressive balance effect: perception and physiological responses of prosody and gestures. *J. Lang. Soc. Psychol.* 41, 659–684. doi:10.1177/0261927x221078317

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

Rohrer, P. L., Vilà-Giménez, I., Florit-Pons, J., Gurrado, G., Gibert, N. E., Ren, A., et al. (2020). "The MultiModal MultiDimensional (M3D) labeling system," in Proceedings of the Gesture and Speech in Interaction (GESPIN), Stockholm, Sweden, September 2020. doi:10.17605/OSF.IO/ANKDX

Rosenberg, A., and Hirschberg, J. (2009). Charisma perception from text and speech. *Speech Commun.* 51, 640–655. doi:10.1016/j.specom.2008.11.001

Rosenthal, R. (1976). *Experimenter effects in behavioral research*. USA: Irvington Publishers, Inc.

Sakib, M. N., Chaspari, T., and Behzadan, A. (2019). "Coupling virtual reality and physiological markers to improve public speaking performance," in Proceedings of the19th International Conference on Construction Applications of Virtual Reality (CONVR2019), Bangkok, Thailand, November 2019, 171–180.

Sanchez-Vives, M., and Slater, M. (2005). From presence to consciousness through virtual reality. *Nat. Rev. Neurosci.* 6, 332–339. doi:10.1038/nrn1651

Scheidel, T. (1967). *Persuasive speaking*. Glenview: Scott Foresman.

Schneider, J., Borner, D., van Rosmalen, P., and Specht, M. (2017). Presentation trainer: what experts and computers can tell about your nonverbal communication. *J. Comput. Assisted Learn.* 33 (2), 164–177. doi:10.1111/jcal.12175

Selck, K., Albert, T., and Niebuhr, O. (2022). "And miles to go before–Are speech intensity levels adjusted to VR communication distances?," in *Book of abstracts of the 13th nordic prosody conference* (Denmark: Sønderborg), 44–46.

Shapiro, L. (2014). *The routledge handbook of embodied cognition*. New York, NY: Routledge.

Siegert, I., and Niebuhr, O. (2021). Case report: women, be aware that your vocal charisma can dwindle in remote meetings. *Front. Commun.* 5, 611555. doi:10.3389/fcomm.2020.611555

Signorello, R., D'Errico, F., Poggi, I., and Demolin, D. (2012). "How charisma is perceived from speech: A multidimensional approach," in Proceedings of the 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing, Amsterdam, Netherlands, September 2012, 435–440. doi:10.1109/SocialCom-PASSAT.2012.68

Slater, M., Pertaub, D., Barker, C., and Clark, D. (2006). An experimental study on fear of public speaking using a virtual environment. *Cyberpsychology Behav. impact Internet, multimedia virtual Real. Behav. Soc.* 9 (5), 627–633. doi:10.1089/cpb.2006.9.627

Slater, M., and Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Front. Robot. AI* 3, 74. doi:10.3389/frobt.2016.00074

Slater, M., Steed, A., McCarthy, J., and Maringelli, F. (1998). The influence of body movement on subjective presence in virtual environments. *Hum. Factors* 40 (3), 469–477. doi:10.1518/001872098779591368

Slater, M., Usoh, M., and Steed, A. (1995). Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Trans. Computer-Human Interact. (TOCHI)* 2, 201–219. doi:10.1145/210079.210084

Smith, C. D., Sawyer, C. R., and Behnke, R. R. (2005). Physical symptoms of discomfort associated with worry about giving a public speech. *Commun. Rep.* 18 (1-2), 31–41. doi:10.1080/08934210500084206

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 (1), 87–100. doi:10.1111/flan.12381

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 (5), e0216288. doi:10.1371/journal.pone.0216288

Tanner, B. A. (2012). Validity of global physical and emotional SUDS. *Appl. Psychophysiol. Biofeedback* 37 (1), 31–34. doi:10.1007/s10484-011-9174-x

Thomas, L. E., and Lleras, A. (2009). Swinging into thought: directed movement guides insight in problem solving. *Psychonomic Bull. Rev.* 16, 719–723. doi:10.3758/PBR.16.4.719

Thrasher, T. (2022). The impact of virtual reality on L2 French learners' language anxiety and oral comprehensibility: an exploratory study. *CALICO J.* 39 (2), 219–238. doi:10.1558/cj.42198

Tse, A. Y. (2012). Glossophobia in university students of Malaysia. *Int. J. Asian Soc. Sci.* 2 (11), 2061–2073.

Vallade, J. I., Kaufmann, R., Frisby, B. N., and Martin, J. (2020). Technology acceptance model: investigating students' intentions toward adoption of immersive 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. (2022). Unguided virtual-reality training can enhance the oral presentation skills of high-school students. *Front. Commun.* 7, 910952. doi:10.3389/fcomm.2022910952

Valls-Ratés, I. (2023). *Using virtual reality to train public speaking skills in a secondary school setting*. Doctoral dissertation. Barcelona, Spain: Universitat Pompeu Fabra.

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. Comput. Assisted Learn.* 36, 412–422. doi:10.1111/jcal.12424

Wagner, U., Gais, S., Haider, H., Verleger, R., and Born, J. (2004). Sleep inspires insight. *Nature* 427 (6972), 352–355. doi:10.1038/nature02223

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 (3), 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. *Israel J. psychiatry Relat. Sci.* 48 (2), 91–97.

Wang, F., Lee, E. K. O., Wu, T., Benson, H., Fricchione, G., Wang, W., et al. (2014). The effects of tai chi on depression, anxiety, and psychological well-being: A systematic review and meta-analysis. *Int. J. Behav. Med.* 21, 605–617. doi:10.1007/s12529-013-9351-9

Weber, M. (1968). *On charisma and institution building*. Chicago, IL, USA: University of Chicago Press.

Weninger, F., Krajewski, J., Batliner, A., and Schuller, B. (2012). The voice of leadership: models and performances of automatic analysis in online speeches. *IEEE Trans. Affect. Comput.* 3 (4), 496–508. doi:10.1109/t-affc.2012.15

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bull. Rev.* 9 (4), 625–636. doi:10.3758/BF03196322

Wolpe, J. (1969). "Pergamon general psychology series," in *The practice of behavior Therapy* (Elmsford, NY, US: Pergamon Press).

Xu, Y. (2013). "ProsodyPro - a tool for large-scale systematic prosody analysis," in Proceedings of the Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), Aix-en-Provence, France, January 2013, 7–10.

Yokoyama, H., and Daibo, I. (2012). Effects of gaze and speech rate on receivers' evaluations of persuasive speech. *Psychol. Rep.* 110 (2), 663–676. doi:10.2466/07.11.21.28.PR0.110.2.663-676

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 Psychology/Temas em Psicologia* 27 (2), 491–507. doi:10.9788/tp2019.2-14

Zellin, M., Mühlenen, A., von Müller, H. J., and Conci, M. (2014). Long-term adaptation to change in implicit contextual learning. *Psychon. Bull. Rev.* 21, 1073–1079. doi:10.3758/s13423-013-0568-z