ORIGINAL ARTICLE

Acoustic characteristics of voice production in virtual reality-simulated and physical environments: a comparative study in university professors

Daniel Rodríguez[1,2] · Adrián Borrego[2] · Marco Guzmán[3] · Roberto Llorens[2]

Abstract

This study investigated the reliability of a virtual reality-simulated classroom to generate a comparable self-perception of voice quality and acoustic effects of phonation to a real classroom in a group of teachers, and sense of presence. Thirty university professors participated in the study and were required to produce loud connected speech by reading a 100-word text in two conditions: (1) in a real classroom including a group of students, and (2) in a virtual replica of the classroom consisting of a 360-degree video of the same classroom and students, which was displayed using a head mounted display. Ambient noise was controlled in both conditions by playing classroom noise through headphones. The self-perception of voice quality, the long-term average spectrum and smooth cepstral peak prominence were estimated in both conditions. The sense of presence generated by virtual reality was measured after interacting with the virtual classroom. There were no statistically significant differences in the self-perception of voice quality or in the acoustic measures of voice production between conditions. The sense of presence in the virtual classroom was high. Our findings suggest that a virtual reality-simulated classroom generate comparable self-perception of voice quality and acoustic effects of phonation to the real classroom, and a high sense of presence, in a group of teachers. Additionally, it is important to highlight the potential of virtual reality to enhance the ecological validity of acoustic assessment of voice production in laboratories and clinical settings.

Keywords Virtual reality · Voice assessment · Acoustic analysis · Ecological validity

1 Introduction

Voice disorders, also referred as dysphonia, are very common in teachers, due to their work activity. While the prevalence of dysphonia in the general population have been reported to range from 6 to 15%, the prevalence in teachers can reach between 20 and 50% (Mattiske et al. 1998; Roy et al. 2004). Compared to the general population, teachers show more symptoms of dysphonia, such as hoarseness or vocal fatigue, and have nearly 6 times more work absences related to these disorders (Behlau et al. 2014; Castillo et al. 2015; Roy et al. 2004). Several studies have investigated how teachers use their voice in the classroom setting, concluding that they are more prone to develop vocal pathologies due to prolonged use of their voice (Vilkman 2000), poor working conditions, such as increased background noise, poor climate conditions, and absence of rest (Cantor-Cutiva et al. 2013; Durup et al. 2017; Kankare et al. 2011; Rantala et al. 2015; Vilkman 2000), low vocal hygiene awareness (Bolbol et al. 2017) and lack of vocal training (Ilomäki et al. 2005). These characteristics lead to symptoms associated with vocal discomfort and fatigue (Kankare et al. 2011). Teachers in noisy classrooms tend to increase vocal intensity, leading to augmented phonatory effort and self-perceived tense voice (Abo-Hasseba et al. 2017; Hunter et al. 2020; Laukkanen et al. 2008; Laukkanen and Kankare 2006; Phadke et al. 2019; Rantala et al. 1998; Södersten et al. 2002).

✉ Roberto Llorens
rllorens@htech.upv.es

Daniel Rodríguez
drodriguez@uct.cl

1 Department of Therapeutic Processes, Universidad Católica de Temuco, Campus San Francisco – Temuco, Manuel Montt 56, Temuco, Araucanía, Chile

2 Neurorehabilitation and Brain Research Group, Institute for Human-Centered Technology Research, Universitat Politècnica de València, Valencia, Spain

3 Department of Communication Sciences and Disorders, Universidad de los Andes, Santiago, Chile

The alterations of voice production in response to the environmental conditions and potential emotional consequences associated with dysphonia, including mood and anxiety disorders (Da Rocha and De Mattos Souza 2013) support the need for an accurate assessment and treatment of voice production in teachers. Voice evaluation is conducted by speech and language pathologists, as well as otorhinolaryngologists and other related professionals, to determine the causes and consequences of the vocal disorder. Assessments usually include functional examination using auditory-perceptual voice examination that can be complemented with interviews and self-report questionnaires, acoustic voice analysis, aerodynamic measurements of voice, laryngeal imaging examination and electroglottographic measurements, which all provide complementary information for a comprehensive assessment of voice production (Roy et al. 2013). Although a great part of the most common assessment procedures have shown acceptable validity and reliability and are easy to administer and not time-consuming, the majority of the functional assessment instruments that attempt to measure vocal behavior in daily life, such as the Voice Handicap Index (Jacobson et al. 1997), the Voice-related Quality of Life (Hogikyan and Sethuraman 1999) or the Voice Symptom Scale (Deary et al. 2003), are based on self-reported data (Roy et al. 2013), which can sometimes have limited accuracy and be biased (Chang and Karnell 2004). Interestingly, subjective voice tests and objective laboratory tests can show different results (Hanschmann et al. 2011; Hummel et al. 2010; Woisard et al. 2007). While objective voice assessment instruments can overcome these limitations, they are almost exclusively used (when available) in laboratories or in the clinical setting, which may not adequately match the factors of the environments that originate the dysphonia (Bottalico et al. 2018), e.g. a classroom in the case of teachers, thus limiting the ecological validity of the assessment. Portable dosimeters, equipped with microphones and accelerometers attached to the neck, allow for measuring the fundamental frequency and intensity of the voice with the goal of estimating vocal load in any environment (Popolo et al. 2002). However, their accuracy has been shown to be limited (Bottalico et al. 2018).

Virtual reality (VR) enables the simulation of an environment or activity through real-time multisensory stimulation and real-time user interaction and exploration (Bermúdez i Badia et al. 2016), which facilitates the sense of presence or 'being there' in a virtual environment that replaces the physical world (Lombard and Ditton 1997). VR has the ability to recreate safe, ecologically valid, and individualized environments while registering and objectively measuring behavior and performance within the virtual world. These characteristics have motivated the use of VR in clinical, affective, and social neurosciences (Navarro et al. 2013; Parsons 2015). Although these features have the potential to enable investigation of

voice production in ecological valid environments, the application of VR to speech and voice therapy is scant (Bryant et al. 2020). An immediate improvement in voice measurements has been detected in individuals with Parkinson's disease using VR (Cruz et al. 2020) and its use has also demonstrated positive effects and reduced anxiety for subjects who speak in public (Takac et al. 2019). Remacle et al. (2021) investigated the vocal behavior of a group of 30 female primary school teachers while talking to an experimenter, while teaching in a real classroom, and while teaching in an immersive virtual classroom, which consisted of 16 virtual non-realistic students animated with typical childlike actions, displayed through a head mounted display. Their results indicated that the participants significantly increased the frequency, intensity, and duration of their voice pauses while teaching, either in a real or a virtual classroom, in comparison to during the conversation with the experimenter. Although these findings provide preliminary evidence of the potential of virtual environments to produce comparable demands to real environments, the voice recordings were made in an environment with ambient noise and the lessons in the real and in the virtual classroom were improvised, so vocal emission was not controlled. These limitations therefore prevent analyzing the voice spectrum or voice disturbances isolated from the room noise and performing long-term average spectrum analysis, cepstral analysis or perturbation measurements, which require adequate and comparable recording conditions.

Despite the limitations, this preliminary report highlighted the potential of VR to generate ecologically valid environments, such as classrooms for teachers, to investigate voice production and objectively measure vocal behavior. Additionally, virtual classrooms have been shown to elicit a satisfactory sense of presence (Reeves et al. 2021; Takac et al. 2019). The purpose of present study was: (1) to investigate the reliability of a virtual classroom to generate comparable acoustic characteristics of voice and self-perceived voice quality to a real classroom in a group of teachers; and (2) to investigate the sense of presence perceived by the teachers while being in the virtual classroom. We hypothesized that VR can generate a plausible and immersive ecologically valid classroom that evokes comparable demands of voice production to real environments, which would be reflected in comparable acoustic characteristics of voice and high sense of presence.

2 Method

2.1 Participants

A convenience sample of teachers was recruited from the faculty of Health School of the Catholic University of Temuco. The participation criteria were, first, to work as a teacher; second, to have a teaching load of 10 h or

more per week in classes that require vocal use; and third, to show at least one symptom of dysphonia on the Voice Symptom Scale (Deary et al. 2003).

Thirty university professors, 19 woman and 11 men, participated in this study. Participants had a mean age of 37.1 ± 6.6 years, which ranged from 28 to 52 years. Participants had an average teaching experience of 9.5 ± 4.5 years of teaching, using loud voice for an average of 12.7 ± 5.8 h per week. Participants also reported that they use their voice at conversational intensity for an average of 32.0 ± 21.3 h per week. Participants showed an average of 21.8 symptoms of dysphonia, ranging from 5 to 42, in the Voice Symptom Scale (Table 1).

The study was accepted by the ethics committee of Catholic University of Temuco ID N°J1-10928. All participants gave their written informed consent prior to their participation in the study.

2.2 Instrumentation

A Focusrite Scarlett 2i4 USB audio interface (Focusrite PLA Inc., High Wycombe, UK) and an Audiotechnica model AT2020 omnidirectional condenser microphone (Audiotechnica Inc., Tokio, Japan) were used to record audio samples during the experiment and for the preparation of the auditory stimuli. The recorded signals were digitized at a sampling rate of 44.1 kHz and 16 bits. The Praat software (Boersma and Weenink 2022), version 6.2.14, was used for all recordings.

A 5 min 360-degree video of a university classroom with 8 students talking to each other was recorded using a GoPro Max camera (GoPro Inc., San Mateo, CA, USA) with a 4 K resolution (3840×2160) and 60 fps. The camera was placed on a pedestal at a height of 170 cm (5′7″), simulating the average height of the participants, with its back to the blackboard pointing towards the students. The microphone was placed 15 cm in front of the participant's mouth, and the pedestal with the text was placed on the left side of the image at 30 cm. Participants remained standing during the experiment. The text was phonetically balanced and consisted of 100 words (Martínez-Cifuentes et al. 2020), which took approximately 60 s to read. Finally, a dedicated laptop and an audio interface, which was connected to the microphone, were placed on a remote desk. The students sat on the left side of the classroom (on the right side from the participants' point of view) as the left side was mostly hidden by the poster board with the text (Fig. 1). The noise of the room was recorded using the microphone.

The recorded video was displayed through a VR head-mounted display, the Oculus Quest (Meta Platforms Inc., Menlo Park, CA, USA) with a video resolution of 4 K.

Table 1 Individual scores in the voice symptom scale

Participant	Functional [0–60]	Emotional [0–32]	Physical [0–28]	Total [0–120]
1	14	1	6	21
2	9	2	3	14
3	5	2	13	20
4	6	1	3	10
5	9	1	9	20
6	13	3	4	21
7	13	0	6	19
8	14	1	6	23
9	2	0	5	7
10	19	2	16	38
11	17	0	17	34
12	15	0	7	22
13	10	0	5	15
14	0	0	1	1
15	25	0	16	42
16	12	0	14	26
17	23	1	17	41
18	14	1	13	28
19	11	0	5	16
20	12	1	10	24
21	16	0	12	28
22	17	2	9	29
23	11	0	3	14
24	3	0	8	11
25	7	0	4	11
26	21	4	9	36
27	10	4	9	24
28	5	0	15	20
29	12	4	4	21
30	0	0	3	4

Individual scores in the Functional, Emotional and Physical subscales and total score of the Voice Symptom Scale

Auditory stimuli were played by a smartphone and provided by semi in-ear headphones, the Samsung EG920 (Samsung Inc., Seul, South Corea), with a flat 20 Hz–20 kHz frequency response. These headphones allow for some sound leakage around the ear canal, enabling the wearer to receive auditory feedback while speaking.

2.3 Procedure

The experiment took place in the same classroom that was used to record the 360-video. The room had no acoustic treatment and the noise was kept below 45 dB SPL in all conditions. An experimenter conducted and supervised all the sessions.
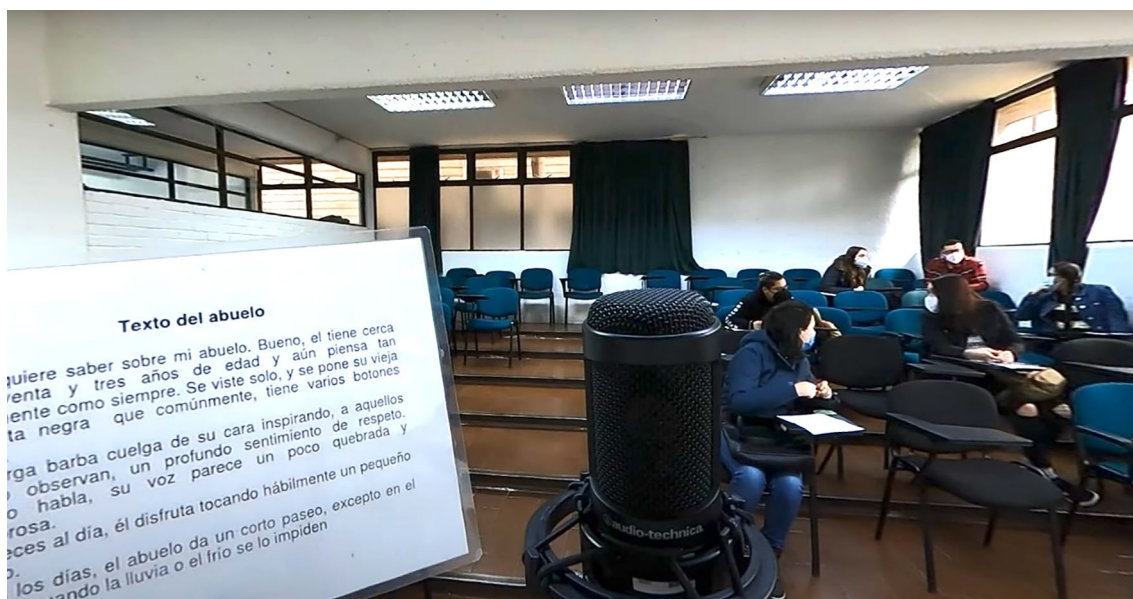
Fig. 1 Snapshot of the recorded 360-video from the participant's point of view

Prior to the experiment, the participants were informed about the objective of the study, signed the informed consent, provided information about their age, sex, years of teaching, hours using a projected voice per week, hours using a conversational voice per week, and completed the Voice Symptom Scale. Then, the experiment started.

Participants were required to stand in the platform of the classroom, located where the camera was positioned during the recording of the 360 video, and to read the same text mentioned above from the poster board by projecting their voice under two different conditions, a virtual classroom (*VR*) and a real classroom *(in-person)*, administered in counterbalanced order (Fig. 2). In the *VR condition*, participants were equipped with the VR headset, which displayed the recorded video of the real classroom, and read the text

directly from the video. In the *in-person condition*, participants performed the same procedure but in the real world. The position of the elements (microphone, poster board with the text, etc.) and the same 8 students who appeared in the recording video remained the same in both conditions. The students in the *in-person condition* simulated their actions in the recorded video without making a sound, reproducing similar mouth and hand movements but in silence. In both conditions, a background conversation noise obtained while recording the 360 video was played through the headphones at an average of 60 dB SPL.

Consequently, in both conditions the participants saw the microphone on a pedestal, the poster board with a printed text on another pedestal, and the students talking while hearing them murmuring.
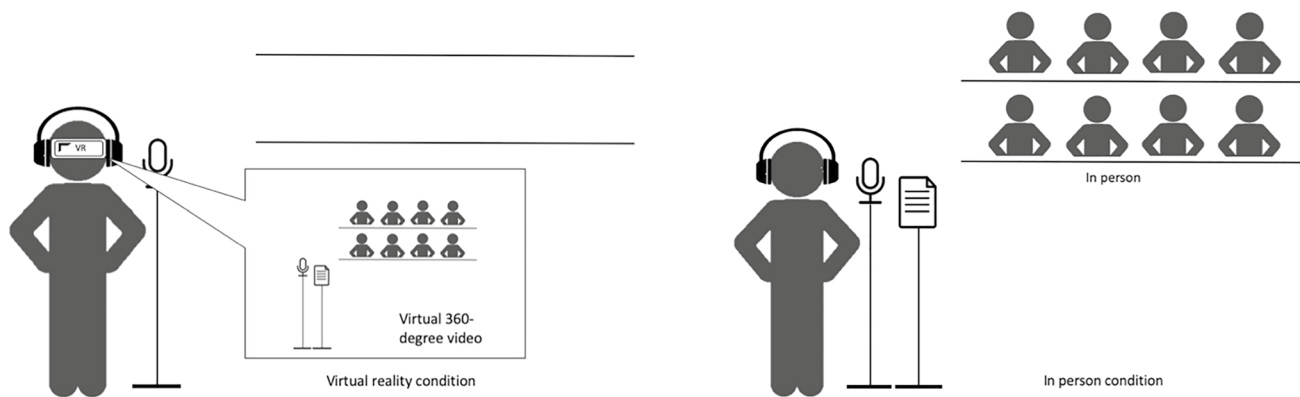

Fig. 2 Experimental setting. Virtual reality condition (left) and In person condition (right)

After reading the text in each condition, participants were required to self-assess their voice quality. A 100-mm visual analogue scale was used, where 0 indicated low voice quality and 100 indicated high voice quality. After the *VR condition* participants additionally assessed the level of presence elicited by the experience in the Slater-Usoh-Steed questionnaire (Usoh et al. 2000) and a shortened version of the Presence Questionnaire (Witmer and Singer 1998), which only included items 4, 5, 7, 8, 10, 12, 14, 15, 16, 19, 21, 23, 24, 25, and 26 of the original questionnaire. Both instruments included items rated on a 7-point Likert scale, where values ranging from 1 to 3 were considered as indicators of absence of presence, values of 4 were considered as neutral, and values 5 to 7 were considered as indicators of high sense of presence.

Acoustic analysis of the voice recorded during the text reading was performed with the Long-Term Average Spectrum, which provides information on the average energy of the voice frequency spectrum range (Löfqvist and Mandersson 1987), and Cepstral Peak Prominence Smoothed (CPPS), which represents the harmonic quality of the voice.

3 Data analysis

3.1 Acoustic analysis

A 100 Hz bandwidth and a Hanning window were used for each recording. Prior to the Long-Term Average Spectrum analysis, the non-speech audio portions were removed from the audio recordings.

The Long-Term Average Spectrum analysis included (1) the L1-L0 ratio, which represents the difference in sound pressure between the F1 (300–800 Hz) and the F0 (50–300 Hz); (2) the alpha ratio, which represents the sound level difference between 50–1000 Hz and 1000–5000 Hz; and (3) the 1/5–5/8 ratio, which represents the sound level difference between 1000–5000 Hz and 5000–8000 Hz. The L1-L0 ratio has been associated with the degree of glottic adduction. Hypoadduced voices have been shown to have a strong L0 (or strong sound level of F0) and a low L1 (low sound level of F1), while voices with high glottic adduction have been shown to have a weak L0 and a strong L1 (Kitzing 1986). The Alpha Ratio represents the general spectral curve, which has been shown to depend on the type of

phonation (degree of vocal fold adduction), being higher in hyperfunctional voices (Laukkanen et al. 2008). The 1/5–5/8 ratio has been associated to the level of breathiness noise in the voice, being lower in people with less breathiness in phonation (Hammarberg et al. 1980).

With regards to the CPPS analysis, high values have been associated with a voice with less loss of spectral energy and absence of dysphonia (Maryn and Weenink 2015; Murton et al. 2020). On the contrary, low values have been associated with worse spectral energy, worse vocal quality and a higher degree of dysphonia.

All acoustic analysis was conducted using Praat software, version 6.2.14.

3.2 Statistical analysis

Normality of the data was tested with the Shapiro Wilks test. All the measures but the L1-L0 showed a normal distribution, thus parametric statistics were used.

Differences in the measures of self-assessed voice quality and acoustic analysis obtained in the *VR condition* and *in-person condition* were investigated using Student's t-tests.

The α level was set at 0.05 for all analyses (two-sided). Statistical analyses were performed using SPSS version 25 (IBM, Armonk, NY, USA).

4 Results

4.1 Voice quality

Participants rated their vocal quality in the *in-person condition* with a mean of 76.6 ± 15.8 over 100, and in the *VR condition* of 71.2 ± 15.9. No statistical differences were found between conditions in the self-assessment of voice quality ($t(29) = 2.00$, $p = 0.054$), although there was a trend towards significance.

4.2 Acoustic measures

No statistically significant differences were found in the measurements of the L1-L0 ratio, the alpha ratio, the 1/5–5/8 ratio, or the CPPS between conditions (Table 2).

Table 2 Acoustic measures obtained during the virtual reality condition and the in person condition

Measure	Virtual reality condition	In-person condition	Significance
L1-L0 ratio (dB SPL)	1.38 ± 2.69,	1.67 ± 3.02	$t(29) = 0.79$, $p = 0.43$
Alpha ratio (dB SPL)	$-15.60 \pm 2{,}64$	-15.80 ± 2.69	$t(29) = -0.77$, $p = 0.44$
1/5–5/8 ratio (dB SPL)	-15.86 ± 3.48	-16.07 ± 3.60	$t(29) = -0.62$, $p = 0.53$
CPPS (dB SPL)	8.42 ± 0.73	8.41 ± 0.87	$t(29) = -0.10$, $p = 0.91$

4.3 Sense of presence

Participants experienced a high sense of presence in the virtual classroom, as reflected by scores of 5.6 ± 1.2 over 7 and 5.28 ± 1.3 over 7 in both the Slater-Usoh-Steed questionnaire and the modified version of the Presence Questionnaire, respectively.

Results in the Presence Questionnaire showed that auditory presence (questions 4, 14 and 15 of the original instrument) achieved the highest score with an average score of 6.1 ± 0.8 over 7, and visual presence (questions 5, 10 and 12) was rated with an average score of 5.56 ± 1.3 over 7. These values proved that audiovisual stimulation was successful at delivering a reliable perception of being in the virtual classroom, and was also supported by self-reports of having been inside the virtual world (question 16), which was rated with an average score of 5.80 ± 1.3 over 7, and having been immersed in the experience (question 23), which was rated with an average score of 5.66 ± 1.3 over 7. Participants also reported to have been focused on the reading task (question 25), with an average score of 5.96 ± 1.1 over 7, and not being distracted during the experience (question 24), with an average score of 2.1 ± 1.3 over 7.

5 Discussion

This study investigated the reliability of a VR-simulated classroom to generate acoustic effects of phonation in comparison to a real classroom and sense of presence in a group of teachers. Our results indicated that teachers had comparable self-perception of voice quality and comparable acoustic measures, including the L1-L0 ratio, the alpha ratio, the 1/5–5/8 ratio, and the CPPS, and a high sense of presence in the virtual classroom. These findings evidence the potential of VR to generate ecologically valid environments that could allow for objective and accurate acoustic assessment of voice production during real-life activities in laboratories and clinical settings.

Our results of self-reported voice quality are supported by previous studies that also used a visual analog scale to investigate this construct in subjects with functional dysphonia, which reported values that ranged from 75 to 80 (Frisancho et al. 2020; Guzman et al. 2017), consistent with our findings. Although the self-reports of voice quality in both conditions were comparable, there was a tendency towards significance that might suggest a tendency of the participants to feel higher voice quality in the *in-person condition* compared to the *VR condition*. In spite of the fact that good correlations have been found between self-perceived voice features measured using a visual analogue scale and acoustic measures of the voice (Castillo-Allendes et al. 2022), self-perceived measures of voice features have been suggested as

being less reliable than acoustic measures (Park and Stepp 2019).

The results in the acoustic measures of our study are analogously supported by previous studies with similar objectives and procedures. First, the values of the L1-L0 ratio found in our study are in line with a previous study by Master et al (2008), which reported increasing values of the measure with the strength of the vocal intensity. Specifically, authors found L1-L0 values of 0.45 for conversational vocal intensity, 2.8 for moderate intensity, and 3.3 for strong vocal intensity. Our results are consistent with those values, and suggest that participants used a moderate vocal intensity during the text reading in both environments. Interestingly, previous research has showed an increase in the L1-L0 ratio after phonatory effort in teachers (Laukkanen et al. 2004), which has also been correlated with a low level of voice breathiness (Laukkanen et al. 2001, 2004). Second, the values of the alpha ratio found in our study are comparable to those found by Rantala et al. (2015) in teachers. The authors found a mean value of -15.20 ± 2.06 in the alpha ratio values measured with a text reading sample after teaching classes in noisy environments, such as noise from electronic devices together with the noise produced by students in a classroom. In their study, Rantala et al. also concluded that the alpha ratio decreased as more vocal effort was demanded, and argued that this effect can be associated with a hypokinetic voice, probably related to the vocal fatigue caused by the vocal load after a day of teaching. It is expected that this effect is limited or does not even occur in less noisy environments, where vocal production becomes more hyperfunctional (Laukkanen et al. 2008). Third, the value of the 1/5–5/8 ratio found in our study is consistent with that of a previous study that investigated the emission of voice in anger (Guzman et al. 2013). The authors found an average value of the 1/5–5/8 ratio of -13.04 ± 3.80, which is very similar to our findings. The use of both an angry voice and a projected voice might require a comparable increased intensity and decreased air output, in comparison to a neutral conversational voice. Finally, some studies have shown that the use of a projected voice in the classroom is reflected by increased CPPS values. The studies by Maryn and Weenink (2015) and Phadke et al (2020) reported comparable CPPS values of 11.66 ± 2.68 and 11.4 ± 1.4, respectively, which are higher than that found in our study. The differences in the results might be related to a different distance to the microphone (15 cm in our study and 6 cm in previous studies). It is important to consider that CPPS values can be influenced by the type of phonatory task. Higher CPPS values have been reported at higher vocal intensity (Brockmann-Bauser et al 2021), which could be a relevant factor to consider in future research that compare different phonatory tasks.

The comparable acoustic measures of voice production in the virtual classroom and the real classroom in our study are

supported by Remacle et al (2021), who reported that giving a lecture of a self-preferred topic to computer-generated virtual students elicited similar vocal emissions to those in a real classroom. In contrast to their study, our experiment required participants to read a text, which controlled for the type of vocal emission, used headphones to provide auditory stimulation, which controlled for the noise level, and used real students who replicated the same behavior in both conditions, which controlled for the visual stimulation. These conditions allowed for a reliable comparison between conditions and, additionally, analyzing the Long-Term Average Spectrum, which enabled investigating the vocal intensity used in the different frequencies of the vocal spectrum and not only in the fundamental frequency. The absence of differences in the measures of the Long-Term Average Spectrum and the CPPS in both the *VR* and *in-person* condition, which is supported by the comparable self-perceived voice quality in both conditions, evidence that VR can potentially simulate ecologically valid environments that require comparable demands of voice production to real classrooms.

This is particularly relevant for the assessment of voice production in case of occupational dysphonia, as the conditions that require increased vocal demands are not commonly present in the environment where the assessment is conducted, which potentially limits the extrapolation of the findings and the validity of the measures (González-Gamboa et al. 2022). Franzen and Wilhelm (1996) defined ecological validity according to truthfulness, the extent to which the performance on the assessments predicts the performance on the activities of interest; and plausibility, the extent to which measures have similar requirements to those required in the activities of interest. Furthermore, Zaki and Ochsner (2009) investigated critical differences between laboratory and real life situations and suggested that ecologically valid assessments should include multisensory stimuli (visual, auditory, linguistic, etc.) that must be provided as they occur in the activity of interest, and contextual environments that allow for emotional interpretation of the situation. VR can simulate real-life activities and provide controlled real time audiovisual stimuli consistent with real-life environments, which grants the ability to generate ecologically valid environments. This feature has motivated the use of VR to facilitate the clinical assessment of abilities or behaviors that are associated to specific contexts (Navarro et al. 2013; Parsons 2015). A recent study by Daşdöğen et al (2023) investigated the effects of visual, auditory, and audiovisual stimulation in VR classrooms varying in noise and size. Their findings revealed that multimodal stimulation, combining visual and auditory stimulation, elicited distinct vocal effects compared to individual modalities, with participants exhibiting decreased vocal loudness and effort, and increased vocal comfort. These findings support the use of multimodal

stimulation to achieve ecologically valid contextualization of environments relevant to voice users.

The high values of presence found in this study are in line with the results of vocal behavior, as participants reported to have experienced a strong sense of being in the virtual classroom, which could have motivated a comparable behavior to that in the real classroom. These results are supported by previous studies that investigated the sense of presence in teachers while exploring 360-videos of a classroom through a head-mounted display (Ferdig and Kosko 2020; Gandolfi et al. 2021). All these findings support the ability of immersive videos to elicit a successful sense of being in a classroom with the advantage of not having to be physically in the environment.

Although this study highlights the potential application of VR to the assessment of voice production, future applications of this technology should consider the attitude of speech therapists towards the use of VR in clinical settings. Vaezipur et al. (2022) reported that speech therapists have a positive attitude towards this type of technologies. In their study, speech therapists valued the ability of VR to generate realistic environments where patients can train speech and language skills. Importantly, Bryant et al (2022) discusses the ethical implications of using VR. The authors highlight the need to generate virtual environments that recreate real scenarios that respect the disability situation through empathy. Along these lines, the authors suggest that VR applications should be co-designed by patients and specialized clinical professionals to reduce the risks of unethical designs. For example, in the area of voice, Smith et al (2023) concluded that the use of VR would allow for the creation of safe environments for voice training, where people can make mistakes with less negative consequences. Brassel et al (2023) delves into the ethical aspects associated with the use of VR in individuals with functional impairments after a traumatic brain injury, and identify several factors that may limit the use of VR in people with visual, cognitive, sensory or physical disabilities. Additionally, the authors suggest that the design of VR applications should consider minimizing side effects such as dizziness or anxiety. All these challenges should be addressed to facilitate the clinical integration of VR into the clinical practice of speech therapists.

The limitations of this study should be taken into account when analyzing the results. First, the characteristics of the class, such as the number of students, the ambient noise, and the projected voice phonatory task, may not be representative of other contexts and, consequently, extrapolation of the results should be done with caution. Second, all the participants had at least one symptom of dysphonia. Consequently, the results of the experiment in people without any voice impairment is unknown. Third, participants did not embody a virtual avatar in the virtual classroom, which some participants perceived as being "floating". Although this is

expected to limit the sense of embodiment and presence in the virtual environment (Ventura et al. 2022), our results evidenced a high sense of presence. The effect that this limitation may have on vocal production is unknown. Fourth, the interaction with the virtual environment was limited, as the recorded 360 video did not allow for interactive communication with the students, and the text to read was not very representative of a teaching lesson, as it corresponded to a standard text used for voice evaluation. Fifth, although reading a standardized text helped to generate comparable audio data among participants, it does not accurately represent the task of giving a lecture, which might involve more complex cognitive mechanisms. This limitation should be considered when extrapolating the results. Finally, we did not control for the attention driven towards the text or to the rest of the environment. Consequently, the effects of attention to different parts of the environment on the sense of presence and vocal production is unknown. This could have affected the subjects' perception of the activity, and therefore their experience within the experiment.

However, our findings suggest that a 360-video of a classroom can generate comparable self-perceived voice quality and acoustic phonation effects to a real class, together with high levels of presence in the virtual classroom. This reveals the potential of VR to improve the ecological validity of voice assessment, which is conventionally performed through self-reports that can be biased and have limited accuracy. The acoustic analysis of voice production is commonly restricted to dedicated laboratories and clinical settings, which might fail to simulate the conditions where the voice is challenged. VR can overcome the difficulties at conducting instrumented assessment of voice production in real environments, by simulating the conditions of real environments in the laboratory or clinical setting with high levels of presence. This enables investigating a series of variables that are difficult to measure in uncontrolled environments, such as the spectral curve, the spectrogram, long-term average spectrum measurements, the use of formants in speech, measures derived from cepstrum, among others. This analysis can enhance the assessment of ecological vocal use and the degree of functional alterations in the daily life of people with dysphonia. In conclusion, the combined used of acoustic measures of voice production and interaction in ecologically valid environments could improve the accuracy and objectiveness of voice assessment.

6 Conclusions

A VR-simulated classroom, consisting of a recorded 360 video of a real classroom, could generate comparable self-perception of voice quality and acoustic effects of phonation to the real classroom, and a high sense of presence, in a group of teachers. These findings highlight the potential of VR to improve the ecological validity of acoustic assessment of voice production in laboratories and clinical settings.

Author contributions Conceptualization—DR, RL, AB; Methodology—DR, RL, AB, MG; Formal analysis and investigation—DR, RL; Writing—original draft preparation—DR; Writing—review and editing: RL, MG; Funding acquisition—DR; Supervision—RL, AB.

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

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.

Ethical approval The study was accepted by the ethics committee of Catholic University of Temuco ID N°J1-10928. All participants gave their written informed consent prior to their participation in the study.

References

Abo-Hasseba A, Waaramaa T, Alku P, Geneid A (2017) Difference in voice problems and noise reports between teachers of public and private schools in upper Egypt. J Voice 31:508.e11-508.e16. https://doi.org/10.1016/j.jvoice.2016.10.016

Behlau M, Zambon F, Madazio G (2014) Managing dysphonia in occupational voice users. Curr Opin Otolaryngol Head Neck Surg 22:188–194. https://doi.org/10.1097/MOO.0000000000000047

Bermúdez i Badia S, Fluet GG, Llorens R, Deutsch JE (2016) Virtual reality for sensorimotor rehabilitation post stroke: design principles and evidence. Neurorehabilitation technology, 2nd edn. Springer, Cham, pp 573–603

Boersma P, Weenink D (2022) Praat

Bolbol SA, Zalat MM, Hammam RAM, Elnakeb NL (2017) Risk factors of voice disorders and impact of vocal hygiene awareness program among teachers in public schools in Egypt. J Voice 31:251.e9-251.e16. https://doi.org/10.1016/j.jvoice.2016.07.010

Bottalico P, Ipsaro Passione I, Astolfi A et al (2018) Accuracy of the quantities measured by four vocal dosimeters and its uncertainty. J Acoust Soc Am 143:1591–1602

Brassel S, Brunner M, Power E et al (2023) Speech-language pathologists' views of using virtual reality for managing cognitive-communication disorders following traumatic brain injury. Am J Speech Lang Pathol 32:907–923

Brockmann-Bauser M, Van Stan JH, Sampaio MC et al (2021) Effects of vocal intensity and fundamental frequency on cepstral peak prominence in patients with voice disorders and vocally healthy controls. J Voice 35:411–417

Bryant L, Brunner M, Hemsley B (2020) A review of virtual reality technologies in the field of communication disability: implications for practice and research. Disabil Rehabil Assist Technol 15:365–372. https://doi.org/10.1080/17483107.2018.1549276

Bryant L, Sedlarevic N, Stubbs P et al (2022) Collaborative co-design and evaluation of an immersive virtual reality application prototype for communication rehabilitation (DISCOVR prototype). Disabil Rehabil Assist Technol 19:1–10

Cantor-Cutiva LC, Vogel I, Burdorf A (2013) Voice disorders in teachers and their associations with work-related factors: a systematic review. J Commun Disord 46:143–155. https://doi.org/10.1016/j.jcomdis.2013.01.001

Castillo A, Casanova C, Valenzuela D, Castañón S (2015) Prevalencia de disfonía en profesores de colegios de la comuna de santiago y factores de riesgo asociados. Cienc Trab 17:15–21. https://doi.org/10.4067/s0718-24492015000100004

Castillo-Allendes A, Guzmán-Ferrada D, Hunter EJ, Fuentes-López E (2022) Tracking occupational voice state with a visual analog scale: voice quality, vocal fatigue, and effort. Laryngoscope 133(7):1676–1682

Chang A, Karnell MP (2004) Perceived phonatory effort and phonation threshold pressure across a prolonged voice loading task: a study of vocal fatigue. J Voice 18:454–466. https://doi.org/10.1016/j.jvoice.2004.01.004

Cruz T, Coriolano M, Silva H et al (2020) The immediate effect of vocal technique associated with virtual reality game in people with Parkinsons disease. Eur J Public Health 30:040–006

Da Rocha LM, De Mattos Souza LD (2013) Voice handicap index associated with common mental disorders in elementary school teachers. J Voice 27:595–602. https://doi.org/10.1016/j.jvoice.2012.10.001

Daşdöğen Ü, Awan SN, Bottalico P et al (2023) The influence of multisensory input on voice perception and production using immersive virtual reality. J Voice. https://doi.org/10.1016/j.jvoice.2023.07.026

Deary IJ, Wilson JA, Carding PN, MacKenzie K (2003) VoiSS: a patient-derived voice symptom scale. J Psychosom Res 54:483–489. https://doi.org/10.1016/S0022-3999(02)00469-5

Durup N, Shield BM, Dance S, Sullivan R (2017) Teachers' voice parameters and classroom acoustics—a field study and online survey. J Acoust Soc Am 141:3540–3540. https://doi.org/10.1121/1.4987482

Ferdig RE, Kosko KW (2020) Implementing 360 video to increase immersion, perceptual capacity, and teacher noticing. TechTrends 64:849–859

Franzen MD, Wilhelm KL (1996) Conceptual foundations of ecological validity in neuropsychological assessment. Ecological validity of neuropsychological testing. Gr Press/St Lucie Press Inc, Grand Rapids, pp 91–112

Frisancho K, Salfate L, Lizana K et al (2020) Immediate effects of the semi-occluded ventilation mask on subjects diagnosed with functional dysphonia and subjects with normal voices. J Voice 34:398–409. https://doi.org/10.1016/j.jvoice.2018.10.004

Gandolfi E, Kosko KW, Ferdig RE (2021) Situating presence within extended reality for teacher training: validation of the extended reality presence scale (XRPS) in preservice teacher use of immersive 360 video. Br J Educ Technol 52:824–841. https://doi.org/10.1111/bjet.13058

González-Gamboa M, Segura-Pujol H, Oyarzún PD, Rojas S (2022) Are occupational voice disorders accurately measured? A systematic review of prevalence and methodologies in schoolteachers to report voice disorders. J Voice. https://doi.org/10.1016/j.jvoice.2022.10.023

Guzman M, Correa S, Muñoz D, Mayerhoff R (2013) Influence on spectral energy distribution of emotional expression. J Voice 27:129.e1-129.e10. https://doi.org/10.1016/j.jvoice.2012.08.008

Guzman M, Jara R, Olavarria C et al (2017) Efficacy of water resistance therapy in subjects diagnosed with behavioral dysphonia: a randomized controlled trial. J Voice 31:385.e1-385.e10. https://doi.org/10.1016/j.jvoice.2016.09.005

Hammarberg B, Fritzell B, Gaufin J et al (1980) Perceptual and acoustic correlates of abnormal voice qualities. Acta Otolaryngol 90:441–451. https://doi.org/10.3109/00016488009131746

Hanschmann H, Lohmann A, Berger R (2011) Comparison of subjective assessment of voice disorders and objective voice measurement. Folia Phoniatr Logop 63:83–87

Hogikyan ND, Sethuraman G (1999) Validation of an instrument to measure voice-related quality of life (V- RQOL). J Voice 13:557–569. https://doi.org/10.1016/S0892-1997(99)80010-1

Hummel C, Scharf M, Schuetzenberger A et al (2010) Objective voice parameters and self-perceived handicap in dysphonia. Folia Phoniatr Logop 62:303–307

Hunter EJ, Cantor-Cutiva LC, van Leer E et al (2020) Toward a consensus description of vocal effort, vocal load, vocal loading, and vocal fatigue. J Speech Lang Hear Res 63:509–532. https://doi.org/10.1044/2019_JSLHR-19-00057

Ilomäki I, Mäki E, Laukkanen AM (2005) Vocal symptoms among teachers with and without voice education. Logop Phoniatr Vocol 30:171–174. https://doi.org/10.1080/14015430500294106

Jacobson BH, Johnson A, Grywalski C et al (1997) The voice handicap index (VHI) development and validation. Am J Speech Lang Pathol 6:66–70

Kankare E, Geneid A, Laukkanen AM, Vilkman E (2011) Subjective evaluation of voice and working conditions and phoniatric examination in kindergarten teachers. Folia Phoniatr Logop 64:12–19. https://doi.org/10.1159/000328643

Kitzing P (1986) LTAS criteria pertinent to the measurement of voice quality. J Phon 14:477–482. https://doi.org/10.1016/s0095-4470(19)30693-x

Laukkanen AM, Kankare E (2006) Vocal loading-related changes in male teachers' voices investigated before and after a working day. Folia Phoniatr Logop 58:229–239. https://doi.org/10.1159/000093180

Laukkanen AM, Syrjä T, Laitala M, Leino T (2004) Effects of two-month vocal exercising with and without spectral biofeedback on student actors' speaking voice. Logop Phoniatr Vocol 29:66–76. https://doi.org/10.1080/14015430410034479

Laukkanen AM, Ilomäki I, Leppänen K, Vilkman E (2008) Acoustic measures and self-reports of vocal fatigue by female teachers. J Voice 22:283–289. https://doi.org/10.1016/j.jvoice.2006.10.001

Laukkanen A, Vintturi J, Vilkman E et al (2001) Perceptual, acoustic and self-reported correlates of vocal loading. In: Proceedings of the 25th World congress of the international association of logopedics and phoniatrics in montreal 5.-9.8. 2001

Löfqvist A, Mandersson B (1987) Long-time average spectrum of speech and voice analysis. Folia Phoniatr Logop 39:221–229. https://doi.org/10.1159/000265863

Lombard M, Ditton T (1997) At the heart of it all: The concept of presence. J Comput Commun 3:JCM321. https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Martínez-Cifuentes R, Torres-Bustos V, Sáez-Carrillo K (2020) Passages used in the assessment of Chilean adults with neurological speech disorders. Rev Logop Foniatr Audiol 40:77–82. https://doi.org/10.1016/j.rlfa.2019.11.002

Maryn Y, Weenink D (2015) Objective dysphonia measures in the program praat: smoothed cepstral peak prominence and acoustic voice quality index. J Voice 29:35–43. https://doi.org/10.1016/j.jvoice.2014.06.015

Master S, De Biase N, Chiari BM, Laukkanen AM (2008) Acoustic and perceptual analyses of Brazilian male actors' and nonactors' voices: long-term average spectrum and the "actor's formant." J Voice 22:146–154. https://doi.org/10.1016/j.jvoice.2006.09.006

Mattiske JA, Oates JM, Greenwood KM (1998) Vocal problems among teachers: a review of prevalence, causes, prevention, and treatment. J Voice 12:489–499

Murton O, Hillman R, Mehta D (2020) Cepstral peak prominence values for clinical voice evaluation. Am J Speech Lang Pathol 29:1596–1607

Navarro MD, Lloréns R, Noé E et al (2013) Validation of a low-cost virtual reality system for training street-crossing. A comparative study in healthy, neglected and non-neglected stroke individuals. Neuropsychol Rehabil 23:597–618. https://doi.org/10.1080/09602011.2013.806269

Park Y, Stepp CE (2019) Test–retest reliability of relative fundamental frequency and conventional acoustic, aerodynamic, and perceptual measures in individuals with healthy voices. J Speech Lang Hear Res 62:1707–1718

Parsons TD (2015) Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. Front Hum Neurosci 9:660

Phadke KV, Abo-Hasseba A, Švec JG, Geneid A (2019) Influence of noise resulting from the location and conditions of classrooms and schools in upper Egypt on teachers' voices. J Voice 33:802.e1-802.e9. https://doi.org/10.1016/j.jvoice.2018.03.003

Phadke KV, Laukkanen AM, Ilomäki I et al (2020) Cepstral and perceptual investigations in female teachers with functionally healthy voice. J Voice 34:485.e33-485.e43. https://doi.org/10.1016/j.jvoice.2018.09.010

Popolo PS, Rogge-Miller K, Svec JG, Titze IR (2002) Technical considerations in the design of a wearable voice dosimeter. J Acoust Soc Am 112:2304

Rantala L, Paavola L, Körkkö P, Vilkman E (1998) Working-day effects on the spectral characteristics of teaching voice. Folia Phoniatr Logop 50:205–211. https://doi.org/10.1159/000021462

Rantala L, Hakala S, Holmqvist S, Sala E (2015) Classroom noise and teachers' voice production. J Speech Lang Hear Res 58:1397–1406. https://doi.org/10.1044/2015_JSLHR-S-14-0248

Reeves R, Elliott A, Curran D et al (2021) 360° video virtual reality exposure therapy for public speaking anxiety: a randomized controlled trial. J Anxiety Disord 83:102451. https://doi.org/10.1016/j.janxdis.2021.102451

Remacle A, Bouchard S, Etienne AM et al (2021) A virtual classroom can elicit teachers' speech characteristics: evidence from acoustic measurements during in vivo and in virtuo lessons, compared to a free speech control situation. Virtual Real 25:935–944. https://doi.org/10.1007/s10055-020-00491-1

Roy N, Merrill RM, Thibeault S et al (2004) Prevalence of voice disorders in teachers and the general population. J Speech Lang Hear Res 47:281–293. https://doi.org/10.1044/1092-4388(2004/023)

Roy N, Barkmeier-Kraemer J, Eadie T et al (2013) Evidence-based clinical voice assessment: a systematic review. Am J Speech Lang Pathol 22:212–226. https://doi.org/10.1044/1058-0360(2012/12-0014)

Smith C, Gregory C, Bryant L (2023) Utilizing virtual reality for gender-affirming voice training: surveying the attitudes and perspectives of potential consumers. Int J Lang Commun Disord. https://doi.org/10.1111/1460-6984.12968

Södersten M, Granqvist S, Hammarberg B, Szabo A (2002) Vocal behavior and vocal loading factors for preschool teachers at work studied with binaural DAT recordings. J Voice 16:356–371. https://doi.org/10.1016/S0892-1997(02)00107-8

Takac M, Collett J, Blom KJ et al (2019) Public speaking anxiety decreases within repeated virtual reality training sessions. PLoS ONE 14:e0216288. https://doi.org/10.1371/journal.pone.0216288

Usoh M, Catena E, Arman S, Slater M (2000) Using presence questionnaires in reality. Presence 9:497–503

Vaezipour A, Aldridge D, Koenig S et al (2022) "It's really exciting to think where it could go": a mixed-method investigation of clinician acceptance, barriers and enablers of virtual reality technology in communication rehabilitation. Disabil Rehabil 44:3946–3958

Ventura S, Cebolla A, Latorre J et al (2022) The benchmark framework and exploratory study to investigate the feasibility of 360-degree video-based virtual reality to induce a full body illusion. Virtual Real. https://doi.org/10.1007/s10055-021-00567-6

Vilkman E (2000) Voice problems at work: a challenge for occupational safety and health arrangement. Folia Phoniatr Logop 52:120–125. https://doi.org/10.1159/000021519

Witmer BG, Singer MJ (1998) Measuring presence in virtual environments: a presence questionnaire. Presence 7:225–240

Woisard V, Bodin S, Yardeni E, Puech M (2007) The voice handicap index: correlation between subjective patient response and quantitative assessment of voice. J Voice 21:623–631

Zaki J, Ochsner K (2009) The need for a cognitive neuroscience of naturalistic social cognition. Ann N Y Acad Sci 1167:16–30