Teachers’ speech in a virtual classroom

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

Angélique Remacle, PhD¹,²,³, Stéphane Bouchard, PhD⁴,⁵, Anne-Marie Etienne, PhD², Marie-Chr\-istine Rivard⁴, Dominique Morsomme, PhD²

¹ Fund for Scientific Research – F.R.S.-FNRS, Brussels, Belgium
² Département de Logopédie, Faculté de Psychologie, Logopédie et Sciences de l’Éducation,
  Université de Liège, Liège, Belgium
³ Faculté des Sciences Psychologiques et de l’Éducation, Université Libre de Bruxelles, Brussels,
  Belgium
⁴ Département de Psychoéducation et de psychologie, Université du Québec en Outaouais,
  Gatineau, Canada
⁵ Centre de Santé et de Services Sociaux de l’Outaouais, Gatineau, Canada

Corresponding author: Dr. Angélique Remacle

Mailing address: Université de Liège
Faculté de Psychologie, Logopédie et Sciences de l’Education
Département de Logopédie
Unité Logopédie de la Voix
Rue de l’Aunaie, 30 (B38)
4000 Liège (Belgium)

Phone number: +32 4 366 36 02
Email address: angelique.remacl galerie@uliege.be

This is a post-peer-review, pre-copyedit version of an article published in Virtual Reality.
The final authenticated version is available online at: http://dx.doi.org/10.1007/s10055-020-00491-1
Teachers’ speech in a virtual classroom

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

Abstract

To achieve pedagogic goals and deal with environmental constraints such as noise when lecturing, teachers adapt their speech production in terms of frequency, intensity, and temporal aspects. The mastery of appropriate vocal skills is key to teachers’ speech intelligibility, health, and educational effectiveness. This project tests the relevance of virtual reality (VR) for training teachers’ vocal skills by simulating a lesson in a realistic VR environment characterized by adjustable constraints such as background noise and fidgety children. The VR environment depicts an elementary school classroom with 16 pupils aged 9 to 12 years old animated with typical childlike actions.

To validate this virtual classroom in terms of speech characteristics, we conducted acoustic analyses on the speech productions of 30 female teachers in three conditions: (1) giving a free speech while facing the experimenter (control), (2) teaching in their usual classroom (in vivo), and (3) teaching the same lesson in a virtual classroom (in virtuo). The background noise in the VR setting was adjusted for each talker so it was similar to the level measured in vivo.

Repeated measures ANOVAs showed that teachers significantly increased their voice frequency, intensity, and intonation, and made longer pauses while speaking in vivo and in virtuo, compared to the control condition (p < .001). These voice and speech adaptations (partly related to background noise), the strong feeling of presence and the lack of side effects suggest that the virtual classroom may facilitate voice training and rehabilitation for teachers.

This is a post-peer-review, pre-copyedit version of an article published in Virtual Reality.
The final authenticated version is available online at: http://dx.doi.org/10.1007/s10055-020-00491-1
1. Introduction

In elementary school, information is predominantly presented orally to children. Teachers use their voices to pursue different goals: (1) pedagogic: the challenge of catching the pupils’ attention, and helping them understand and retain a message (Manfredi and Dejonckere 2016); and (2) classroom management: organization and discipline. Compared to normal speech, teaching represents a specific vocal demand (Hunter et al. 2020) and triggers the intentionality of the voice. Manfredi and Dejonckere (2016) defined the intentionality of the voice as “the need to ‘operate’ on the listener and to reach a goal.” As a vocal demand response (Hunter et al. 2020), teachers adapt their voice production, resulting in modifications of the acoustic voice signal in terms of frequency, intensity, and temporal aspects.

From the acoustic perspective, the teaching voice is characterized by an increase in fundamental frequency ($f_0$), intensity (Rantala et al. 2015; Remacle et al. 2014), and dynamic range (Manfredi and Dejonckere 2016), or intonation. The standard deviation of $f_0$ ($f_0$ SD) is an indicator of intonation: the greater the standard deviation, the more lively and interesting to listen to the voice is (Hincks 2004). Regarding the temporal aspects of speech, teaching involves turn-taking: the teacher leaves silent pauses (i.e., interruption of the acoustic voice signal) to let the pupils express themselves. By helping to structure and highlight certain parts of a discourse, pauses help listeners to better integrate the message (Béchet et al. 2013). For the speaker, pauses make it possible to plan what will be said: the more cognitively demanding a task is, the higher the rate and duration of pauses (Kreiman and Sidtis 2011).

In addition, a teacher’s voice is subject to constraints related to the audience (e.g., a group of pupils), the speaking mode (e.g., voice projection), and the environment, such as distance to the interlocutors and background noise impeding speech intelligibility (Schiller et al. In press). Background noise triggers automatic adaptation by the speaker, who involuntarily increases...
Teachers’ speech in a virtual classroom

his/her vocal intensity and frequency to improve audibility: this is called the Lombard effect (Lombard 1911). To be more intelligible in background noise or when communicating with children, speakers decrease their speech rate and hyperarticulate, resulting in formant frequency changes (Smiljanic and Bradlow 2009). While \( f_o \) depends on the sound source (i.e., frequency of vocal fold oscillation), formant frequencies also depend on resonance of the supraglottal vocal tract and influence the sound spectrum (Titze et al. 2015). Regarding voice production, increased intensity and \( f_o \) are associated with a greater vocal load (Rantala et al. 2015; Remacle et al. 2014; Schiller et al. 2018) and vocal effort (Phadke et al. 2019), which can lead to phonotrauma and pathologies such as vocal fold nodules and polyps (Manfredi and Dejonckere 2016). At some point in their career, more than half of all teachers experience a voice disorder (Van Houtte et al. 2011). Because vocal behavior plays an important role in the pathogenesis of such disorders (Manfredi and Dejonckere 2016), the mastery of appropriate vocal skills is key to teachers’ health, and their educational effectiveness.

Although they can improve a teacher’s voice, prevention and treatment programs do not include the training of vocal skills in everyday teaching situations (Faham et al. 2016; Hazlett et al. 2011; López et al. 2017; Nanjundeswaran et al. 2012; Pizolato et al. 2013; Richter et al. 2016; Timmermans et al. 2012). In a clinical setting, it is difficult to practice vocal techniques for handling constraints such as speaking to a group of children against background noise. Consequently, it is challenging to generalize efficient speech patterns from the clinic to educational situations involving realistic constraints and maintain them there.

In the field of speech and language pathology, research examining virtual reality (VR) to enable improved communication for people with communication disabilities is limited (Bryant et al. 2019). The need to develop VR environments to provide therapeutic activities has been emphasized (Theodoros 2008). Their value is in simulating real-world, everyday communicative
Teachers’ speech in a virtual classroom

situations to allow specific competences to be practiced, by sparking the patient’s motivation. Research has provided preliminary support for the idea that VR may be an effective tool in communication interventions (Bryant et al. 2019) for individuals with autism (Halabi et al. 2017), aphasia (Marshall et al. 2016), and stuttering (Brundage and Hancock, 2015). To date, the applicability of VR for voice training and rehabilitation has not yet been studied.

This study fits into the general approach that considers VR to be a tool for training teachers’ vocal skills by simulating a communicative situation (a lesson) in a realistic environment (a classroom), characterized by environmental constraints that are often encountered (background noise, fidgety children). To evaluate the ecological validity of a virtual classroom in terms of vocal behavior in a teaching situation, we studied a group of teachers in three conditions: (1) speaking in a control condition, (2) teaching in their usual classroom (in vivo), and (3) teaching the same lesson in a virtual classroom (in virtuo).

Compared to normal speech, the response to teaching vocal demand should trigger the acoustic indices of Lombard speech, vocal intentionality (Manfredi and Dejonckere 2016) and effort. Thus, in the in vivo and in virtuo conditions, voices would be expected to have a higher $f_0$ (H1), greater intensity (H2), and more intonation contrasts, measured via $f_0$ SD (H3). Teaching involves adaptation of the temporal aspects of speech to promote understanding of the message and interactions with pupils. We would therefore expect more frequent (H4) and longer (H5) pauses in the two classroom conditions than in the control condition.

We will also descriptively document the extent to which the virtual classroom creates a feeling of presence and does not generate unwanted negative side effects (commonly referred to as cybersickness). Finally, the link between measures related to VR exposure and vocal adaptation in the virtual classroom will be examined.
2. Methods

This study was conducted in the Department of Speech Therapy of the University of Liège (Belgium). The Ethics Committee of the Faculty of Psychology, Speech Therapy, and Education Sciences approved it. All participants provided written informed consent after receiving a complete description of the study. During the recruitment procedure, they completed the sociodemographic and baseline questionnaires.

2.1. Participants

The sample included 30 female elementary school teachers aged 22 to 55 years old (M = 50; SD = 10). Their teaching experience ranged between 1 and 35 years (M = 14; SD = 9). Inclusion criteria: be a native French speaker; teach grade 4, 5 or 6 (pupils aged 9 to 12 years old) in a school in the French community in Belgium; be a woman. Exclusion criteria: report a voice or hearing disorder at the time of the study; report a strong tendency to motion sickness.

2.2. Procedure

The experiment took place on two consecutive days. On the first day, the teacher gave a typical lesson of her choice to her pupils in her usual classroom (in vivo condition). The average background noise level was measured in dB(A) (LAEq) throughout the lesson, using a calibrated Rion NL-21 sound level meter (Japan), placed at the back of the classroom at the height of the teacher’s head. The noise level (including both ambient noise and teacher’s speech) measured in vivo ranged from 54.6 to 74.7 dB LAeq (M = 63.5; SD = 4) across the 30 classrooms.

The next day at the same time, the teacher performed two tasks at the laboratory, in a quiet room (background noise level = 35 dB(A)), measuring 4 x 4 x 2.5 m. The first task was a 5-minute free speech as in Astolfi et al. (2015), which involved introducing herself and describing a normal day, while facing the experimenter from 2 m away (control condition). This condition
Teachers’ speech in a virtual classroom

corresponds to what is usually done by speech pathologists to conduct clinical work with their patients. Second, the participant repeated the same lesson as the previous day in a virtual classroom developed for this research project (in virtuo condition). The background noise in the virtual environment was adjusted for each teacher so it was similar to the level measured in vivo the previous day.

In all three conditions, the teacher’s speech was recorded with a Marantz PMD661MKIII digital recorder (Cumberland, RI) and an AKG C520L head-worn microphone (Vienna, Austria) positioned 5 cm from the mouth. The audio signal was recorded in .wav format, with a sampling frequency of 44.1 kHz and 16-bit resolution.

### 2.3. Virtual environment and equipment

A virtual classroom used in our research lab was adapted for the in virtuo condition. The user wore an Oculus Rift™ head-mounted display.

The environment was an elementary school classroom with 16 pupils aged 9 to 12 years old animated with typical childlike actions (Figure 1). The auditory stimuli were spatialized and played through the speakers of the head-mounted display; they corresponded to background noise without understandable semantic content, as described in Klatte et al. (2010) and Phadke et al. (2019). The experimenter can adjust the volume of three noise sources and the children’s agitation level (Figure 2), which allows for clinical flexibility. Thus, it was possible to standardize our experiment based on the situation each teacher experienced in vivo. For each participant, the background noise generated in the headset speakers was calibrated so that it would be identical to the average noise level (LAeq) measured the previous day in vivo. The agitation level was always set at 30%, corresponding to a slightly restless class.

This is a post-peer-review, pre-copyedit version of an article published in *Virtual Reality*. The final authenticated version is available online at: http://dx.doi.org/10.1007/s10055-020-00491-1
Before the immersive experience, each participant received the following instructions: “You are going to teach your lesson to a class of fourth/fifth/sixth graders [cf. the teacher’s actual grade level]. This lesson will last x minutes [same duration as in vivo]. Please try to give the same lesson as yesterday. The virtual environment doesn’t allow for interaction: if you ask questions, the pupils won’t answer you.”

2.4. Measures

2.4.1. Baseline descriptive measures

In addition to sociodemographic information, the following questionnaires were completed.

Voice problems. The Voice Handicap Index (VHI) measures the biopsychosocial impact of voice problems (Jacobson et al. 1997; Woisard et al. 2004). This questionnaire comprises 30 self-rated items on a 5-point Likert scale (0 = never; 4 = always). The total score ranges from 0 (no complaints) to 120 (many complaints).

Immersive tendency. The Immersive Tendencies Questionnaire (ITQ) measures the tendencies of individuals to experience presence in common activities (Witmer and Singer 1998; Robillard et al. 2002). This scale contains 18 self-rated items on a 7-point Likert scale (0 = never; 6 = often).

2.4.2. Measures related to VR exposure

Cybersickness. To control for potential unwanted negative side effects of VR exposure, each participant completed the Simulator Sickness Questionnaire (SSQ) (Bouchard et al. 2009; Kennedy et al. 1993) before and after the in virtuo condition. The SSQ rates 16 symptoms of simulator sickness on a 4-point Likert scale; the scoring is based on raw scores, as recommended by Bouchard et al. (2009).
Teachers’ speech in a virtual classroom

*Presence.* The feeling of presence in VR was measured using the Presence Questionnaire (PQ) (Robillard et al. 2002; Witmer and Singer 1998) completed after the *in virtuo* condition. Nineteen items are self-rated on a 7-point Likert scale, providing scores on six subscales.

### 2.4.3. Voice measures

As mentioned by Smiljanic and Bradlow (2009), there are two broad categories of measurements: (1) global ones including measures such as pause rate and duration, $f_o$ and its variation; and (2) segmental measurements including vowel formant changes, vowel space, and segment duration. As this is the first study on the validity of a virtual classroom for investigating vocal behavior, we opted for global measurements.

For the control, *in vivo* and *in virtuo* recordings of spontaneous productions, acoustic analysis was conducted using Praat freeware (version 6.1.04). The following parameters were extracted from connected speech: median $f_o$ (in Hz) using an autocorrelation method, $f_o$ variation through the standard deviation ($f_o$ SD), and mean voice intensity (in dB). We computed the mean duration (ms) and rate (number per minute) of pauses, considering a minimum silent interval of 350 ms. For each participant, the duration of recordings analyzed was similar for both the *in vivo* and *in virtuo* conditions ($M = 15$ minutes; min–max = 6–24 minutes).

To determine vocal adaptation when teaching in the virtual classroom, we calculated the difference between each voice measure *in virtuo* and during free speech ($\Delta = in virtuo – Control$).

### 2.5. Statistical analysis

To test the hypothesized difference between the teaching and control conditions, repeated measures ANOVAs were conducted with the three conditions as a within-subject factor. The Bonferroni correction was applied considering the five acoustic measures ($\alpha = 0.05/5$). When the $p$-value was lower than 0.01, the null hypothesis of equal means was rejected and simple
orthogonal contrast analyses were performed to test (1) the difference between the control and \textit{in vivo} conditions, and (2) the difference between the control and \textit{in virtuo} conditions. Our aim was to identify voice and speech adaptations in both real and virtual classrooms, the reference point being the free speech condition. The \textit{in vivo} and \textit{in virtuo} conditions were not compared directly for three reasons: (1) comparing all conditions would make the analysis non-orthogonal, thus impacting power and consistency with our research hypotheses; (2) this involves testing the null hypothesis that there is no difference between \textit{in vivo} and \textit{in virtuo} conditions, thus raising the issue of inadequate sample size and power for that specific contrast; and (3) we have no empirical basis to justify a hypothesis that there is, or is not, a difference between \textit{in vivo} and \textit{in virtuo} conditions. Effect sizes are reported using partial $\eta^2$.

The applicability of the analyses was checked in advance, particularly normality using the Shapiro-Wilk test and the assumption of homogeneous variances using Mauchly’s sphericity test. When the sphericity hypothesis was violated, the Greenhouse-Geisser correction adjusting the degrees of freedom is reported. All analyses were performed using SPSS, version 25.

3. Results

3.1. Descriptive results

Regarding baseline descriptive results, the VHI score ($M = 13.83$; $SD = 13.55$) was slightly higher than the normative data for a population without voice disorders ($M = 6.57$; $SD = 5.16$) reported by Woisard et al. (2004). According to these normative data, four teachers exceeded the maximum cut-off score for normality ($=20$) but did not differ from the other participants in their acoustic measures.

Table 1 shows that the participants were not very familiar with VR and video games: 66.7% of them had never tried a VR helmet or glasses, and 73.3% never played video games.
As for VR exposure results, Table 2 reports on the immersive tendency, feeling of presence experienced and unwanted side effects. Qualitatively, the mean ITQ score was within the normal range (Robillard et al. 2002), which means that our participants tended to experience immersion comparable to that of the general population. The PQ scores were within the normal range (Robillard et al. 2002), except for the interface quality subscale: several participants reported that the image was pixelated. Paired t-tests did not show a significant difference in SSQ scores before and after the *in virtuo* condition ($t = .897, p = .377$).

We investigated the potential relationships between sociodemographic variables and voice acoustic measures. In particular, we calculated Pearson correlations between age, teaching experience and each acoustic measure for the control, *in vivo*, and *in virtuo* conditions. As no significant correlation was found, we do not describe these results at length but will focus on the validation of the virtual classroom in terms of vocal behavior.

### 3.2. Voice outcome measures

Table 3 presents the within-subject effect of repeated measures ANOVAs and Figure 3 presents the mean values and SD of the acoustic measures.

Regarding voice frequency, there was a significant effect of condition. The contrast analyses showed higher $f_0$ values *in vivo* ($F(1) = 282.17, p < .001, \eta^2 = .907$) and also *in virtuo* ($F(1) = 202.74, p < .001, \eta^2 = .875$) compared to the free speech control condition produced in the lab. All participants showed a similar adaptation when teaching, consisting in a global $f_0$ rise over the entire lesson *in vivo* and *in virtuo*.

We found a significant effect of condition on voice intensity. Compared to the control condition, contrast analyses indicated that teachers spoke louder *in vivo* ($F(1) = 104.08, p < .001, \eta^2 = .782$) and *in virtuo* ($F(1) = 168.48, p < .001, \eta^2 = .853$). All participants adjusted to the
Teachers’ speech in a virtual classroom

teaching situations by systematically raising their voices, except one who had a similar intensity in the control and in vivo conditions.

Similarly, we found a significant effect of condition on intonation. Compared to free speech, contrast analyses showed greater $f_0$ variations of voice in vivo ($F(1) = 218.76, p < .001, \eta^2 = .883$) and in virtuo ($F(1) = 158.29, p < .001, \eta^2 = .845$). All participants increased their intonation contrasts when teaching in their real classroom. The same adaptation was found in virtuo, except for one teacher who used less intonation when teaching in the virtual classroom than during free speech.

There was no significant effect of condition on the pause rate, meaning that the teachers did not pause more often when teaching compared to the free speech condition. Intersubject variability was observed: 20 participants systematically paused more frequently when teaching (both in vivo and in virtuo) than during free speech; 4 participants systematically paused less frequently when teaching (both in vivo and in virtuo) than during free speech. We also note intrasubject variability of the pause rate: 5 teachers adopted different adaptations across the two teaching conditions.

Finally, we found a significant effect of condition on the duration of pauses. All speakers made longer pauses while teaching in vivo ($F(1) = 106.05, p < .001, \eta^2 = .785$) than during free speech. When teaching in the virtual classroom, 23 out of 30 teachers made longer pauses compared to the free speech condition ($F(1) = 4.25, p = .048, \eta^2 = .128$).

Table 4 presents the relationships between vocal adaptations in the virtual classroom and the global scores on the questionnaires. None of the Pearson correlations was statistically significant, meaning that the vocal changes observed in the virtual classroom are not related to immersive tendencies, the feeling of presence, cybersickness symptoms, or reported voice problems.
4. Discussion

This study aimed to validate the use of a virtual classroom in terms of adaptation of teachers’ speech characteristics in a teaching situation (a lesson) in a realistic environment (a classroom), featuring environmental constraints (background noise, restless pupils), comparing three conditions: (1) free speech produced in a control condition, (2) teaching in vivo, and (3) teaching in virtuo. We wanted to find out whether, compared to the control condition, teaching triggered specific acoustic changes, both in vivo and in a virtual classroom. From a methodological perspective, teachers were instructed to give the same lesson in vivo and in virtuo. The time between the conditions was strictly controlled (1 day). The intensity of background noise was individually controlled for each participant and was similar in both teaching conditions.

The results correspond to acoustic measures of the speech signal in the three conditions. Global measures were calculated on each entire recording of spontaneous speech, and may be complemented by segmental measures to identify some temporal and local variations. In line with H1, compared to speech during the control situation, which represented what is traditionally done in a clinic, the $f_0$ was significantly higher in a teaching situation, both in vivo (+68 Hz or 5.5 semitones on average) and in virtuo (+41 Hz or 3.5 semitones). Similarly, teachers spoke significantly louder when teaching in vivo (+6.1 dB) and in virtuo (+6.3 dB), compared to the control situation, which supports H2. As depicted in Figure 3b, the difference between the in vivo and in virtuo teaching conditions was clinically negligible (+1 dB). These results suggest that the nature of the $f_0$ and intensity adaptations was similar in both teaching situations, but the amplitude of the changes is less pronounced when teaching in the virtual classroom.

Previous studies measured higher voice $f_0$ and intensity in teachers speaking in their professional environment compared to extra-professional situations (Remacle et al. 2014; Schiller
Increased $f_o$ and intensity in the teaching conditions may be triggered by background noise (Rantala et al. 2015; Schiller et al. 2018). Lombard speech has previously been described in noisy environments such as schools (Rantala et al. 2015). This kind of automatic adaptation is likely to provoke increased vocal effort (Phadke et al. 2019) and vocal load (Rantala et al. 2015; Schiller et al. 2018). Professional voice users such as teachers need excellent vocal skills so they can make themselves understood in noisy environments without adopting phonotraumatic behavior. The results of this study show that the virtual classroom seems to allow teachers to practice these skills, as teachers produce specific vocal demand responses (Hunter et al. 2020) even when the stimuli are virtual.

In line with H3, significantly more intonation contrast ($f_o$ variation) was produced when teaching in vivo (+25 Hz) and in virtuo (+17 Hz) than in the control situation. From a qualitative point of view, the participants adopted similar intonation adaptations when teaching, with a less pronounced change in the virtual classroom. This prosodic feature may help teachers to make their speech pleasant and interesting to listen to (Hincks, 2004), to keep the pupils’ attention and fulfill pedagogic goals. In addition, the characteristics of child-directed speech may have applied in the teaching conditions: adults spontaneously adopt higher $f_o$ and greater $f_o$ variability when speaking to children (Saint-Georges et al. 2013).

Contrary to H4, statistical analysis did not indicate that teachers pause more often during lessons, either in vivo or in virtuo. This measure showed particularly large intersubject variability. But in accordance with H5, the duration of pauses was significantly longer when teaching in vivo (+928 ms) and in virtuo (+176 ms), compared to the control condition. Longer pauses during teaching may allow turn-taking, which was difficult to simulate in virtuo given that the environment did not provide for any audience reaction to the user’s speech. In the presence of...
Teachers’ speech in a virtual classroom

background noise, longer pauses may also reflect a slower speech rate with the aim of improving intelligibility (Astolfi et al. 2015).

To sum up, the changes in acoustic parameters were statistically significant, with large effect sizes. Qualitatively speaking, the magnitude of differences compared to the control condition was smaller in virtuo than in vivo. However, the acoustic changes in the VR condition systematically reflected the same kinds of speech adaptation as in a real classroom. In a clinical setting, adjusting the noise level and pupil agitation level in the virtual classroom could make it possible to gradually increase the difficulty of the task and to elicit speech adaptations of similar amplitude to real teaching situations.

In addition to the acoustic measures presented above, the experimenter reported the following subjective observations: while teaching in virtuo, the participants projected their voices, looked at the pupils, moved around the virtual environment, and gestured with their arms to punctuate their speech. For a more general description of communication in VR environments, it would be interesting to use sensors to capture variables related to posture, gestures and eye movements.

Finally, some technical limitations may have affected the VR experience. The first was the lack of interaction with the audience: the virtual characters did not answer questions and did not react to the teacher’s speech. This could have reduced the feeling of co-presence and social presence, although these factors were not measured in this study. Future VR classrooms could benefit from implementing real-time audience feedback, as in VR applications used to train presenters’ skills (El-Yamri et al. 2019; Schneider et al. 2019). Second, teachers reported missing the possibility of using written material (e.g., notes on a sheet or a screen), and of writing on a blackboard. These suggestions represent possible future improvements that could enhance the user experience and the effectiveness of the virtual environment.
5. Conclusions

Learning communication skills such as effective voice use while teaching requires practice in situations that are as similar as possible to reality. VR simulations may represent a promising method to offer several applications for voice training for future teachers (instruction) and rehabilitation for dysphonic individuals (speech and language pathology). In this respect, our study validated the effects of a virtual classroom on the acoustic characteristics of teachers’ speech during a lesson. The results suggest that the virtual classroom elicited changes in voice frequency and intensity, and in speech intonation contrasts and pauses, compared to a control speaking situation. The strong feeling of presence and the lack of side effects such as cybersickness may promote the generalization of communication skills practiced in the virtual classroom to real-world situations. The next step is to conduct randomized controlled trials to assess the efficacy of VR for vocal skill training and rehabilitation.

Acknowledgments

Angélique Remacle was supported by the Fund for Scientific Research – FNRS (F.R.S.-FNRS, Brussels, Belgium). This work was supported by a Canada Research Chairs grant awarded to Stéphane Bouchard and a grant from the Commission Mixte Permanente Québec/Wallonie-Bruxelles awarded to Stéphane Bouchard and Anne-Marie Etienne. We thank Amandine Regnier for her assistance with data collection; Sysmex Belgium N.V. for providing the NL-21 sound level meter used in the study; and Jean-Jacques Embrechts, of the Montefiore Institute at the University of Liège, for his assistance with the calibration of the noise level in the VR environment.
Teachers’ speech in a virtual classroom

Author Disclosure Statement

Stéphane Bouchard is president and part owner of In Virtuo, a company that distributes virtual environments. Conflicts of interest are managed under UQO’s conflicts of interest policy. For all the other authors, no competing financial interests exist.

References


Teachers’ speech in a virtual classroom


This is a post-peer-review, pre-copyedit version of an article published in Virtual Reality. The final authenticated version is available online at: http://dx.doi.org/10.1007/s10055-020-00491-1
Teachers’ speech in a virtual classroom


This is a post-peer-review, pre-copyedit version of an article published in Virtual Reality. The final authenticated version is available online at: http://dx.doi.org/10.1007/s10055-020-00491-1


https://doi.org/10.1016/j.jvoice.2015.05.005


https://doi.org/10.1371/journal.pone.0078103


https://doi.org/10.1016/j.jvoice.2017.06.020
Teachers’ speech in a virtual classroom


This is a post-peer-review, pre-copyedit version of an article published in Virtual Reality. The final authenticated version is available online at: http://dx.doi.org/10.1007/s10055-020-00491-1
Teachers’ speech in a virtual classroom

Teachers’ speech in a virtual classroom

**Figure 1** Virtual environment used for the *in virtuo* condition

This is a post-peer-review, pre-copyedit version of an article published in *Virtual Reality*. The final authenticated version is available online at: http://dx.doi.org/10.1007/s10055-020-00491-1
Teachers’ speech in a virtual classroom

**Figure 2** Control panel for the parameters of the virtual environment. Above: individual adjustment of three sources of background noise: (1) noise from the playground, (2) noise from the corridor, and (3) noise from inside the classroom. Below: adjustment of the pupils’ agitation level from 0 (very attentive and not very restless) to 100 (very distracted and restless). For this study, all three noise sources were adjusted to the same level. For each participant, the noise level played through the head-mounted speakers was identical to the average noise level in the *in vivo* condition. The agitation level was set to 30 for all participants.

**Figure 3** Descriptive results of vocal parameters of teachers giving a free speech in the lab (control), a lesson in their real-world classroom (*in vivo*) and the same lesson while immersed in VR (*in virtuo*), with background noise matched to the level in their *in vivo* classroom. Mean values are reported in the bars, while error bars represent ±1 SD.
Teachers’ speech in a virtual classroom

This is a post-peer-review, pre-copyedit version of an article published in Virtual Reality. The final authenticated version is available online at: http://dx.doi.org/10.1007/s10055-020-00491-1
Teachers’ speech in a virtual classroom

Table 1. Description of the sample of 30 female teachers enrolled in the study

<table>
<thead>
<tr>
<th>Socio-professional variables</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Teaching level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 4</td>
<td>10</td>
<td>33.3</td>
</tr>
<tr>
<td>Grade 5</td>
<td>9</td>
<td>30</td>
</tr>
<tr>
<td>Grade 6</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>Grades 5 and 6</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Grades 4, 5, and 6</td>
<td>2</td>
<td>6.7</td>
</tr>
<tr>
<td><strong>Work schedule</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time</td>
<td>23</td>
<td>76.7</td>
</tr>
<tr>
<td>Part-time</td>
<td>7</td>
<td>23.3</td>
</tr>
<tr>
<td><strong>Voice variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leisure involving voice use</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theater</td>
<td>2</td>
<td>6.7</td>
</tr>
<tr>
<td>Singing in a choir</td>
<td>1</td>
<td>3.3</td>
</tr>
<tr>
<td>Team sport</td>
<td>2</td>
<td>6.7</td>
</tr>
<tr>
<td>None</td>
<td>25</td>
<td>83.3</td>
</tr>
<tr>
<td>Had received voice education during studies or career</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>Had consulted a medical doctor or a speech therapist for a voice problem</td>
<td>4</td>
<td>13.3</td>
</tr>
<tr>
<td>Had consulted a medical doctor for a gastro-esophageal problem</td>
<td>10</td>
<td>33.3</td>
</tr>
<tr>
<td>Smoking</td>
<td>5</td>
<td>16.7</td>
</tr>
<tr>
<td><strong>Digital and virtual reality habits</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often do you play games on your smart phone?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Every day</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Several times a week</td>
<td>5</td>
<td>16.7</td>
</tr>
<tr>
<td>Several times a month</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Less often</td>
<td>16</td>
<td>53.3</td>
</tr>
<tr>
<td>I don’t have a smart phone</td>
<td>1</td>
<td>3.3</td>
</tr>
<tr>
<td>How often do you play video games?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Every day</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Several times a week</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Several times a month</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Less often</td>
<td>5</td>
<td>16.7</td>
</tr>
<tr>
<td>Never</td>
<td>22</td>
<td>73.3</td>
</tr>
<tr>
<td>How comfortable do you feel with digital technologies?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (not at all comfortable)</td>
<td>1</td>
<td>3.3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3.3</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>26.7</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>43.3</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>23.3</td>
</tr>
<tr>
<td>7 (very comfortable)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Have you ever tried a virtual reality helmet or glasses?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>20</td>
<td>66.7</td>
</tr>
<tr>
<td>Less than 3 times</td>
<td>9</td>
<td>30</td>
</tr>
<tr>
<td>From 3 to 5 times</td>
<td>1</td>
<td>3.3</td>
</tr>
</tbody>
</table>

This is a post-peer-review, pre-copyedit version of an article published in *Virtual Reality*. The final authenticated version is available online at: http://dx.doi.org/10.1007/s10055-020-00491-1
Teachers’ speech in a virtual classroom

<table>
<thead>
<tr>
<th>More than 5 times</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

This is a post-peer-review, pre-copyedit version of an article published in *Virtual Reality*. The final authenticated version is available online at: http://dx.doi.org/10.1007/s10055-020-00491-1
Teachers’ speech in a virtual classroom

Table 2. Descriptive results for VR-related variables associated with teaching a lesson while immersed in a virtual classroom

<table>
<thead>
<tr>
<th>Questionnaires</th>
<th>Mean score (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immersive Tendencies Questionnaire (ITQ)(^a)</td>
<td>62.23 (14.45)</td>
</tr>
<tr>
<td>Simulator Sickness Questionnaire (SSQ)</td>
<td></td>
</tr>
<tr>
<td>Before teaching <em>in virtuo</em></td>
<td>5.50 (4.87)</td>
</tr>
<tr>
<td>After teaching <em>in virtuo</em></td>
<td>4.97 (4.54)</td>
</tr>
<tr>
<td>Presence Questionnaire (PQ)(^b)</td>
<td></td>
</tr>
<tr>
<td>Realism</td>
<td>29.6 (8.60)</td>
</tr>
<tr>
<td>Affordance to act</td>
<td>15.17 (5.04)</td>
</tr>
<tr>
<td>Interface quality</td>
<td>8.40 (3.56)</td>
</tr>
<tr>
<td>Affordance to examine</td>
<td>14.13 (3.13)</td>
</tr>
<tr>
<td>Self-evaluation of performance</td>
<td>8.13 (1.59)</td>
</tr>
<tr>
<td>Auditory</td>
<td>16.03 (4.09)</td>
</tr>
</tbody>
</table>

\(^a\). Administered before the experiment  
\(^b\). Administered after the *in virtuo* condition
Teachers’ speech in a virtual classroom

Table 3. Repeated measures ANOVAs comparing vocal parameters of teachers giving a free speech in the lab (control), a lesson in their real-world classroom (*in vivo*) and the same lesson while immersed in VR (*in virtuo*), with background noise matched to the level in the *in vivo* classroom.

<table>
<thead>
<tr>
<th>Parameter Description</th>
<th>df</th>
<th>F</th>
<th>p</th>
<th>η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>f0 median</td>
<td>2, 58</td>
<td>182.02</td>
<td>&lt;.001</td>
<td>.863</td>
</tr>
<tr>
<td>Intensity mean&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.59, 45.99</td>
<td>94.26</td>
<td>&lt;.001</td>
<td>.765</td>
</tr>
<tr>
<td>Intonation (f0, SD)</td>
<td>2, 58</td>
<td>142.36</td>
<td>&lt;.001</td>
<td>.831</td>
</tr>
<tr>
<td>Pause rate&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.53, 44.53</td>
<td>3.31</td>
<td>.058</td>
<td>.102</td>
</tr>
<tr>
<td>Mean duration of pauses</td>
<td>2, 58</td>
<td>56.51</td>
<td>&lt;.001</td>
<td>.661</td>
</tr>
</tbody>
</table>

<sup>a</sup> Greenhouse-Geisser correction reported

This is a post-peer-review, pre-copyedit version of an article published in *Virtual Reality*. The final authenticated version is available online at: http://dx.doi.org/10.1007/s10055-020-00491-1
Teachers’ speech in a virtual classroom

**Table 4.** Correlations between vocal adaptations in the VR classroom (*in virtuo*) and global scores on the questionnaires. For each voice measure, the adaptation when teaching in the VR classroom was obtained by calculating the difference between the *in virtuo* and control conditions ($\Delta = in virtuo – Control$)

<table>
<thead>
<tr>
<th></th>
<th>Voice Handicap Index (VHI)$^a$</th>
<th>Immersive Tendencies Questionnaire (ITQ)$^a$</th>
<th>Simulator Sickness Questionnaire (SSQ)$^b$</th>
<th>Presence Questionnaire (PQ)$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta f_0$ median</td>
<td>$r = -.056, \ p = .77$</td>
<td>$r = -.022, \ p = .91$</td>
<td>$r = .197, \ p = .30$</td>
<td>$r = -.008, \ p = .93$</td>
</tr>
<tr>
<td>$\Delta$ Intensity mean</td>
<td>$r = -.132, \ p = .49$</td>
<td>$r = -.165, \ p = .38$</td>
<td>$r = -.055, \ p = .77$</td>
<td>$r = -.136, \ p = .40$</td>
</tr>
<tr>
<td>$\Delta$ Intonation ($f_0$ SD)</td>
<td>$r = -.129, \ p = .49$</td>
<td>$r = -.108, \ p = .57$</td>
<td>$r = .230, \ p = .22$</td>
<td>$r = -.165, \ p = .33$</td>
</tr>
<tr>
<td>$\Delta$ Pause rate</td>
<td>$r = -.337, \ p = .07$</td>
<td>$r = -.298, \ p = .11$</td>
<td>$r = -.075, \ p = .69$</td>
<td>$r = .111, \ p = .56$</td>
</tr>
<tr>
<td>$\Delta$ Mean duration of pauses</td>
<td>$r = .154, \ p = .42$</td>
<td>$r = -.014, \ p = .94$</td>
<td>$r = .263, \ p = .16$</td>
<td>$r = -.264, \ p = .19$</td>
</tr>
</tbody>
</table>

$a.$ Administered before the experiment  
$b.$ Administered after the *in virtuo* condition

This is a post-peer-review, pre-copyedit version of an article published in *Virtual Reality*.  
The final authenticated version is available online at: http://dx.doi.org/10.1007/s10055-020-00491-1