Towards Virtual Audience Simulation For Speech Therapy

Yann Glémarec
Anne-Gwenn Bosser
Lab-STICC, ENIB
Brest, France

Amelie Hörmann
Norina Lauer
Ostbayerische Technische Hochschule
Regensburg, Germany

Cédric Buche
ENIB, CNRS IRL 2010, CROSSING
Adelaide, Australia

Jean-Luc Lugrin
Marc Erich Latoschik
HCI Group, Informatik IX, University of Wuerzburg
Wuerzburg, Germany

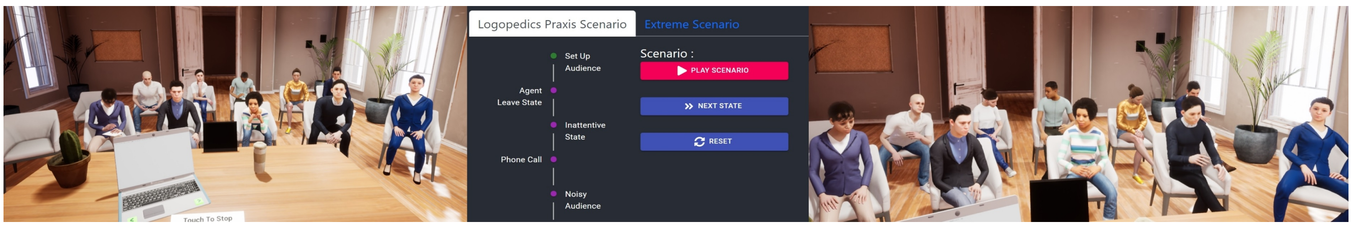

Figure 1: From left to right, the initial state audience's attitude, the therapists' timeline and its controls, and the final scenario state with a distracted and noisy audience.

ABSTRACT

The utilization of virtual reality (VR) technology has shown promise in various therapeutic applications, particularly in exposure therapy for reducing fear of certain situations objects or activities, e.g. fear of height, or negative evaluation of others in social situations. VR has been shown to yield positive outcomes in follow-up studies, and provides a safe and ecological therapeutic environment for therapists and their patients. This paper presents a collaborative effort to develop a VR speech therapy system which simulates a virtual audience for users to practice their public speaking skills. We describe a novel web-based graphical user interface that enables therapists to manage the therapy session using a simple timeline. Lastly, we present the results from a qualitative study with therapists and teachers with functional dysphonia, which highlight the potential of such an application to support and augment the therapists' work and the remaining challenges regarding the design of natural interactions, agent behaviours and scenario customisation for patients.

CCS CONCEPTS

• **Human-centered computing** → *User studies*; **Virtual reality**.

KEYWORDS

datasets, neural networks, gaze detection, text tagging

IVA '23, September 19–22, 2023, Würzburg, Germany

ACM ISBN 978-1-4503-9994-4/23/09.
https://doi.org/10.1145/3570945.3607348

ACM Reference Format:
Yann Glémarec, Anne-Gwenn Bosser, Amelie Hörmann, Norina Lauer, Cédric Buche, Jean-Luc Lugrin, and Marc Erich Latoschik. 2023. Towards Virtual Audience Simulation For Speech Therapy. In *ACM International Conference on Intelligent Virtual Agents (IVA '23), September 19–22, 2023, Würzburg, Germany*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3570945.3607348

1 INTRODUCTION

Functional dysphonia is a voice disorder characterised by hoarseness, breathiness, and vocal fatigue [5]. This disorder can result from various factors, such as stress and anxiety, which can cause tension in the throat's muscles and voice box. Overuse of the voice can also lead to muscle strain and vocal fatigue, which can contribute to functional dysphonia [6]. In some cases, there may be no identifiable cause for the disorder. A speech-language pathologist or voice therapist can evaluate and diagnose functional dysphonia and work with the patient to develop a treatment plan that addresses their individual needs, which may include *transfer work*. This type of voice therapy appears to be effective in treating functional dysphonia [10]. Transfer work typically involves exercises that focus on breath control, relaxation, and proper vocal placement. Patients may also work on exercises that help them to control their vocal volume and pitch, or to use their voice in a way that is efficient and healthy. The objective of transfer work is to help patients achieve a healthy and efficient speaking and singing voice that is free from hoarseness, breathiness, and vocal fatigue.

Virtual reality (VR) has been shown to provide effective results regarding exposure therapy and voice therapy. It shows a significant reduction of anxiety in follow-up studies in situations such as fear of flying, fear of spiders or public speaking anxiety [3, 4, 8]. It also

Figure 2: Silent booth from the Logopädie-Plus practice (A) and the STAGE's virtual environment (B).

shows promise in speech therapy in reducing anxiety over exposure and improving speech performance too [9].

VR technology also provides ecological yet secured environments where the stimulus causing fear can be controlled. In the specific case of public speaking anxiety, VR applications use virtual audiences as fear stimuli and expose the users to various audiences mimicking human behaviours [7].

In this paper, we describe how we defined this new application and the training scenario in collaboration with therapists. The application was used in trials with a cohort of teachers with functional dysphonia. We report the therapist and teachers' feedback regarding the application acceptance gathered through focus group interviews.

2 SPEECH THERAPY APPLICATION

The STAGE is a VR public speaking training application in which users can practice speaking in front of a virtual audience [2]. The virtual audience can reproduce multiple audience attitudes through various non-verbal behaviour, feedback, and context-specific events, e.g. facial expressions, posture, gaze direction or backchannels. The audience's attitude is driven by a Valence-Arousal behaviour model capable of simulating different types of audiences, such as bored, interested, indifferent or enthusiastic [1]. The application also provides the user with an interactive slide presentation system and responsive virtual hands to interact with virtual objects. The virtual hands are animated so when the user is interacting with objects or menus, it is automatically animated accordingly to what and the same with the fingers' location on the VR controller, which uses capacitive sensors to detect their location.

This paper follows from trials of a virtual environment dedicated to speech therapy by therapists with patients with functional dysphonia. Their reports allow to argue in favour of the acceptance of virtual reality audience simulations for therapy exercises, compared to traditional methods relying on patients' imagination. We implemented a 3D replica of the actual practice so that patients would start the session in the virtual place resembling their actual location in the real world: Figure 2, shows the silent booth the patient was using and its virtual representation acting as a passageway between the practice and the therapeutic simulation. To transition between the virtual booth and the VA's virtual training room, when the user triggers a 3D button. The booth then fades out while the user is teleported in the virtuals training room.

3 SUPERVISION APPLICATION AND SCENARIOS

The therapists specified a dedicated scenario based on the VR application features. The development team implemented this scenario using a web-based graphical user interface, tied to a web-based supervision component, which provides controls for the audience's attitude and various high-level events that can be triggered during the simulation, such as background noises or supportive feedback. The encoded scenario then takes the form of a list of events within a timeline that will be triggered accordingly during the simulation. Figure 1 shows the timeline component of the supervision component, as well as the audience displayed before and after the scenario's unfolding. For this study, the scenario's timeline was seven minutes long. After receiving instructions about the study and use of the headsets, patients were isolated into the dedicated real-life booth. After putting on the VR headset, the patients experience the virtual replica of the booth while they get used to the tool. They can then trigger the start of the simulation. The scenario begins with an interested audience which progressively becomes disrespectful and bored until all spectators talk to each other without paying any attention to the speaker, i.e. the current patient. This model characterised the interested attitude as agents smiling, nodding, with opened posture, often leaning forward and rarely gazing away from the speaker. The frequency of these behaviours is also high compare to other attitudes to increase the audience's engagement. Moreover, the model sometime triggers phatic expressions to increase the audience's engagement, i.e. while an agent nods an audio cue such as "mmmhmm" is played out. The scenario calls for a variety of disruptive events to try to unsettle the patient: a spectator comes in late, another leaves, a mobile phone rings loudly. Toward the end of the simulation, a group of agents stops next to the virtual door's room and talks loudly.

4 FEEDBACK FROM END-USERS

The system was tried out by actual therapists and four teachers. The group of study participants was intentionally targeted to teachers. Due to their professional situation, teachers have a particularly high risk of developing a functional voice disorder, e.g. speaking in front of large groups in large rooms and mostly with a lot of background noise. Teachers in Germany do not receive specific voice training in their studies to prevent functional dysphonia. Therefore, teachers are particularly often affected by functional dysphonia and are thus a central target group of speech and language therapy. In the experiment, the professional situation of teachers was simulated. This involves constantly changing demands, so that there are both quieter phases and situations with a high noise level. Teachers must be able to adapt vocally to these situations quickly and use strategies to assert themselves vocally without overloading the voice when noise levels are high.

The qualitative study was divided into three sessions, two individual and one group session. First, teachers practised with standard voice speech exercises and then in a second session they used the VR application before the final focus group interview. Regarding feedback from therapists, the procedure preceding the actual public-speaking simulation is automated, so therapists had no difficulties using the system. From a speech therapist's viewpoint, the system holds potential for therapy in the future. An important advantage of the system from the therapists' perspective was that we implemented a scenario they designed themselves. The simulation was thus easy to use and suited their needs. Therapists also reported that they would use the tool in their everyday practice if it was possible for them to further customise the sessions to their patients' specific needs without technical assistance.

Teachers (patients) reported issues with the VR application which they felt were causing distraction or discomfort during speech therapy exercises in VR, e.g. device's weight, the controllers, and several elements from the virtual environment. Some teachers felt tense during the VR session and reported a strong feeling of immersion. Some agreed that repeated sessions would be beneficial, at least to get used to the device and be more productive. Additionally, setting up the VR equipment was not felt as straightforward, as well as ensuring tracking wasn't lost (which required to start over the exercise). However, teachers reported it is easier to communicate with the virtual audience and support spontaneous reactions.

5 CONCLUSION

We introduced an immersive public-speaking simulation adapted to speech therapy and describe the results of an acceptance study with four teachers with functional dysphonia. The tool is promising: the therapist and the four teachers mentioned they would use it as an intermediate stage after regular therapy sessions. Several issues with the application were mentioned that caused distraction and users reported they could benefit from repeated sessions to get used to the device, overcome technical uneasiness, and become more confident. Therapists reported that casual use would require a future tool to allow easy customisation for the individual needs of patients.

REFERENCES

[1] Yann Glémarec, Jean-Luc Lugrin, Anne-Gwenn Bosser, Aryana Collins Jackson, Cédric Buche, and Marc Erich Latoschik. 2021. Indifferent or Enthusiastic? Virtual Audiences Animation and Perception in Virtual Reality. *Frontiers in Virtual Reality* 2 (2021). https://doi.org/10.3389/frvir.2021.666232

[2] Yann Glémarec, Jean-Luc Lugrin, Anne-Gwenn Bosser, Cédric Buche, and Marc Erich Latoschik. 2022. Controlling the Stage: A High-Level Control System for Virtual Audiences in Virtual Reality. *Frontiers in Virtual Reality* 3 (2022). https://doi.org/10.3389/frvir.2022.876433

[3] Sandra R Harris, Robert L Kemmerling, and Max M North. 2002. Brief virtual reality therapy for public speaking anxiety. *Cyberpsychology & Behavior* 5, 6 (2002), 543–550.

[4] Smiti Kahlon, Philip Lindner, and Tine Nordgreen. 2019. Virtual reality exposure therapy for adolescents with fear of public speaking: a non-randomized feasibility and pilot study. *Child and Adolescent Psychiatry and Mental Health* 13, 1 (2019), 1–10. https://doi.org/10.1186/s13034-019-0307-y

[5] Waltraud Kutej. 2011. *Prävention von Stimmstörungen: die Stimme als wichtiges Arbeitsinstrument in Sprechberufen.* Vol. 8. Schulz-Kirchner Verlag GmbH.

[6] Murray D Morrison, Hamish Nichol, and Linda A Rammage. 1986. Diagnostic criteria in functional dysphonia. *The Laryngoscope* 96, 1 (1986), 1–8.

[7] Maryann E Owens and Deborah C Beidel. 2015. Can virtual reality effectively elicit distress associated with social anxiety disorder? *Journal of Psychopathology and Behavioral Assessment* 37 (2015), 296–305.

[8] Barbara Olasov Rothbaum, Larry Hodges, Samantha Smith, Jeong Hwan Lee, and Larry Price. 2000. A controlled study of virtual reality exposure therapy for the fear of flying. *Journal of Consulting and Clinical Psychology* 68, 6 (12 2000), 1020–1026. https://doi.org/10.1037/0022-006x.68.6.1020

[9] Gareth Walkom. 2016. Virtual Reality Exposure Therapy: To Benefit Those Who Stutter and Treat Social Anxiety. In *2016 International Conference on Interactive Technologies and Games (ITAG)*. 36–41. https://doi.org/10.1109/iTAG.2016.13

[10] Marco Parrino Wiltrud Föcking. 2022. *Praxis der Funktionalen Stimmtherapie.* Vol. 2. Springer Berlin, Heidelberg, Springer Nature, One New York Plaza, Suite 4600, New York, NY, 10004-1562 USA. https://doi.org/10.1007/978-3-662-64579-6