

Public Speaking in Virtual Reality: Facing an Audience of Avatars

Mel Slater,
David-Paul
Pertaub, and
Anthony Steed

University
College London

What happens when someone talks in public to an audience they know to be entirely computer generated—to an audience of avatars? If the virtual audience seems attentive, well-behaved, and interested, if they show positive facial expressions with complimentary actions such as clapping and nodding, does the speaker infer correspondingly positive evaluations of performance and show fewer signs of anxiety? On the other hand, if the audience seems hostile, disinterested, and visibly bored, if they have negative facial expressions and exhibit reactions such as head-shaking, loud yawning, turning away, falling asleep, and walking out, does the speaker infer correspondingly negative evaluations of performance and show more signs of anxiety? We set out to study this question during the summer of 1998. We designed a virtual public speaking scenario, followed by an experimental study.

In this work we wanted mainly to explore the effectiveness of virtual environments (VEs) in psychotherapy for social phobias. Rather than plunge straight in and design a virtual reality therapy tool, we first tackled the question of whether real people's emotional responses are appropriate to the behavior of the virtual people with whom they may interact. We concentrated on public speaking anxiety as an ideal arena for two reasons. First, it's relatively simple technically compared to more general social interactions. A public speaking scenario involves specific stylized behaviors on the part of the avatars, making it relatively straightforward to implement. Second, this application has high usefulness within the broader context of social phobias, public speaking being a prevalent cause of anxiety among the general population.

Previous research into using VEs in a mental health setting have concentrated on specific phobias such as fear of heights, flying, spiders, and open spaces.¹ A recent study suggested that VR might prove useful in treating public speaking anxiety.² This study provided evidence that virtual therapy effectively reduced self-reported levels of anxiety. Our research intends to address the prior step: Before developing an effective tool, we should elicit the factors in a VE that will provoke the desired response in clients. If people do not report and exhibit signs and symptoms similar to those generated during a real public talk, then VE-based therapy cannot succeed. Moreover, the answer to this more

fundamental question can have applications in a wider context than therapy—for example, in our starting point for this research, which took place in the context of collaboration in shared VEs.³

Designing the experiment

The project used DIVE (Distributive Interactive Virtual Environment) as the basis for constructing a working prototype of a virtual public speaking simulation. Developed by the Swedish Institute of Computer Science, DIVE has been used extensively in various national and international research projects investigating the possibilities of VEs. In this multiuser VR system, several networked participants can move about in an artificial 3D shared world and see and interact with objects, processes, and other users present in the world. TCL (Tool Command Language) interpreters attached to objects supply interesting dynamic and interactive behaviors to things in the VE.

We constructed as a Virtual Reality Modeling Language (VRML) model a virtual seminar room that matched the actual seminar room in which subjects completed their various questionnaires and met with the experimenters. (See the sidebar "Technical Details" on page 9 for a synopsis of the equipment we used in our experiment.) The seminar room was populated with an audience of eight avatars seated in a semicircle facing the speaker, as if for a talk held in the real seminar room. These avatars continuously displayed random autonomous behaviors such as twitches, blinks, and nods designed to foster the illusion of "life." We programmed the avatars' dynamic behavior with TCL scripts attached to appropriate body parts in the DIVE database. This let us avoid the situation where only the avatars under the operator's or experimenter's direct control are active while the others stay "frozen" in inanimate poses.

We simulated eye contact by enabling the avatars to look at the speaker. Also, they could move their heads to follow the speaker around the room. Facial animation, based around a linear muscle model developed by Parke and Waters,⁴ allowed the avatars to display six primary facial expressions together with yawns and sleeping faces. Avatars could also stand up, clap, and walk out of the seminar room, cutting across the speaker's line of sight. The avatars could make yawning and clapping

sounds. The accompanying images (Figures 1 through 6) show some of the audience reactions.

An advertisement sent by e-mail to all postgraduate students at University College London invited them to participate in a study that would let them rehearse a short talk (five minutes) in front of a small audience in a safe setting—in VR. We paid five British pounds (about nine US dollars) per subject. Those who agreed to take part completed a questionnaire designed to assess their confidence as public speakers—the Personal Report of Confidence as a Speaker (PRCS). At the end of the study we had full data on 10 subjects. Four of the subjects had a score exceeding 10, which was the average for the group as a whole. A score exceeding 18 indicates a fear of public speaking, so our group had relatively low levels of public speaking anxiety.

The experiment employed a two-factor repeat-measures design. The first factor was immersion, whether subjects gave their talk to the audience displayed on a monitor or “immersed” with a head-mounted display. Each subject repeated their talk three times. The first time they experienced either a very friendly or a very hostile audience reaction. For the second talk, subjects faced whichever audience they did not experience the first time. Whether the audience was “good” or “bad” made up the second factor. The third time the audience always started off with hostile reactions, then switched into very positive reactions. We included this third time for ethical reasons and didn’t use the associated data in the analysis.

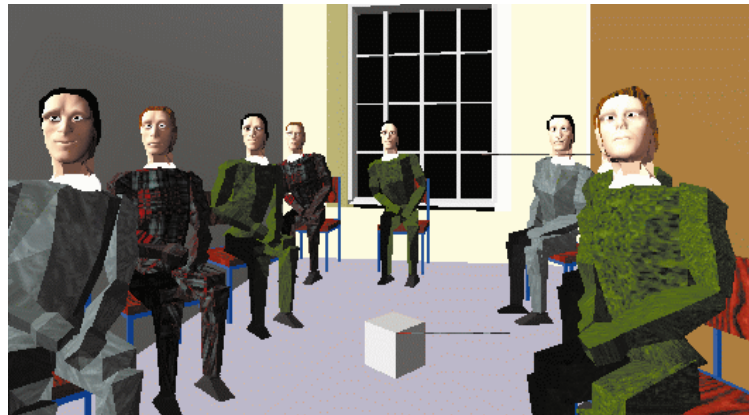
For our experiments, we required the virtual audience to convincingly emote either a pure positive or pure negative response. Audience reactions consisted of stylized animation scripts for individual avatars intended to convey an unambiguous evaluative message. Sequences of these animations formed coherent narratives, identical for all subjects. We devised three such narratives, approximating positive, negative, and mixed audience responses.

We didn’t want entirely to automate audience responses, as speakers would notice if the avatars responded at completely unsuitable points during their talk. We exploited DIVE’s distributed capabilities to allow an unseen operator at a remote workstation to observe the environment as an invisible avatar in the seminar room. The operator could listen to the speech as it unfolded and trigger the next audience response in the current sequence at an appropriate moment. However, only the timing, not the order, of the next audience response was at the discretion of the operator. We did this to equalize the experience across subjects in the experiment.

When subjects arrived, an experimenter took them to the seminar room, explained the procedures to them, and asked them to supply a title for their talk. The experimenter then accompanied the subjects to a nearby VR



1 A receptive “positive” audience greets the speaker with big smiles and lots of applause.



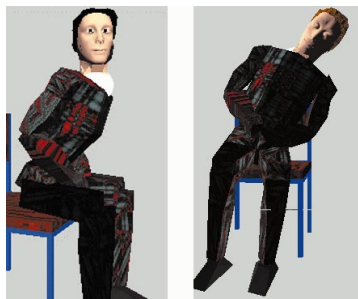
2 This intimidating “negative” audience greets the speaker at the beginning of a presentation.



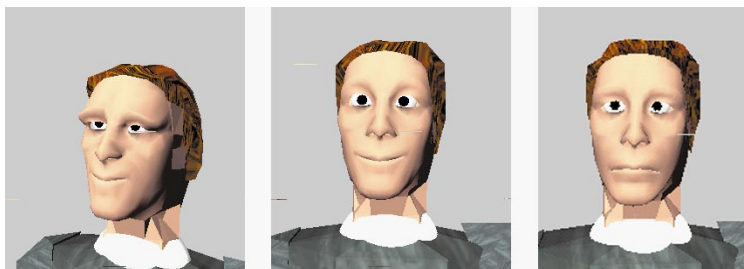
3 They’re listening—but not to the speaker. Members of the audience confer with one another.



4 The man on the left is still listening. But his friend on the right has dozed off.



5 Avatars make faces—disgusted, happy, and sad.



6 A member of the audience walks in front of the speaker on his way out of the room.



laboratory where they were to give their presentation. After each of their three talks, they were taken back to the original room and asked to complete a questionnaire, which related only to the immediately prior talk. They could, however, see and compare their current reactions with their own reactions to previous talks. After all three talks and questionnaires had been completed, subjects were asked to complete a final questionnaire about their confidence in social situations. A short debriefing session then followed, in which they were encouraged to expand on their responses in the questionnaire and discuss their experience of the virtual speaking environment.

Response variable

The main study included three response variables: self-rating of performance, reported physical symptoms of anxiety, and a standard fear of public speaking questionnaire administered after each talk. We only discuss the self-rating set of results here and note that the others are consistent with these. The self-rating question was, "How would you rate your own performance in the talk you have just given? Assign to yourself a score out

of 100, where 0 = completely dissatisfied with your performance and 100 = completely satisfied."

Explanatory variables

In addition to the two main factors (immersion and audience response), we collected data on several potential explanatory variables, including the following.

Background: age and gender. All subjects were in their 20s or 30s, with only one woman among them.

There were seven postgraduate students, one undergraduate, and two faculty members. None of the subjects came from a computer science discipline, and the experimenters knew none of them before the study.

Co-presence. This refers to the sense of being with the virtual audience compared to being with a real audience. Four questions, each on a scale of 1 to 7, elicited information on this response:

- In the last presentation, to what extent did you have a sense that there was an audience in front of you?
- To what extent did you have a sense of giving a talk to people?
- When you think back about your last experience, do you remember this as more like talking to a computer or communicating to an audience?
- To what extent were you aware of the audience in front of you?

Interestingly, we found no significant difference in co-presence scores between the immersed and nonimmersed groups.

Perceived audience interest. Here we attempted to understand the subject's own impressions of audience behavior, which might not match the experimenters' intentions. The question "How would you characterize the interest of the audience in what you had to say?" was scored on a scale of 1 to 7, with higher values indicating higher interest.

Independently of order—whether the "good" or "bad" audience reactions came first—we found a significant difference in perceived audience interest for the two situations: a negative audience (mean 2.5 with standard deviation 1.5) versus a positive audience (mean 4.3 with standard deviation 2.0).

We saw little evidence of a "rehearsal effect," where people's rated performance increases with each successive talk. Although we observed an increase, it proved statistically insignificant. This "time" variable was not statistically significant in any analysis.

Results

This section outlines one of the most important results; a subsequent full paper will give further details.

We assessed the relationship between self-rating and the independent and explanatory variables using nor-

mal multiple regression analysis. The results suggest the following:

- For a negative audience, the higher the perceived audience interest, the higher the self-rating. However, for a positive audience, the perceived interest has no influence on the self-rating.
- For nonimmersed subjects, the rating diminishes with increased co-presence independently of the type of audience response. However, for the immersed subjects, higher co-presence is associated with lower self-rating for the negative audience and higher self-rating for the positive audience.

The regression equations would lead to the following predictions:

- The lowest self-rating would result with a negative audience, immersion, maximum co-presence, and minimum perceived audience interest.
- The highest self-rating would result with alternative combinations:
Negative audience, lowest co-presence, and highest perceived interest.
Positive audience and highest co-presence.

Overall, the regression model provides a very good fit to the data (this model explains 89 percent of the variation in self-rating), and the results seem sensible. We found it noteworthy that when the audience is actually negative, perceived audience interest can overcome the negativity. This result means that the “positive” and “negative” audience responses were not as pure as we aimed for—clearly, sometimes a negative audience reaction was perceived as positive.

We find the results satisfying. In plain language this means that a low self-rating individual immersed in the VE with the virtual audience might say something like, “I felt I was really with these people [high co-presence]. They were behaving terribly [negative audience]. They weren’t at all interested in what I was saying [minimum perceived audience interest].” That’s exactly the kind of response we wanted.

Conclusions

We can conclude

- Higher perceived audience interest increases self-rating and reduces public speaking anxiety.
- Co-presence seems to amplify things, making a “bad” situation worse and a “good” situation better.

A further conclusion important for future studies is that it may not be possible to design “pure” negative or positive audience responses. The perception of the audience response dominates here rather than the value that experimenters place on a particular designed audience response. It’s worth exploring the factors that lead subjects to evaluate an audience as interested or not. Clearly, the actual audience reaction plays a part in this, but it’s not the whole story.

We are treating this study very much as a pilot and

Technical Details

We conducted the experiments on a Silicon Graphics Onyx with twin 196-Mhz R10000 processors, Infinite Reality graphics, and 192 Mbytes of main memory. We used DIVE version 3.3 alpha software, developed at the Swedish Institute of Computer Science. For the immersive sessions, the tracking system had two Polhemus Fastraks, one for the head-mounted display (HMD) and another for a five-button 3D mouse (unused in these experiments). The helmet was a Virtual Research VR4 with a resolution of 742×230 pixels for each eye, 170,660 color elements, and a field of view of 67 degrees diagonal at 85 percent overlap. The frame rate for the experiments varied according to whether the session was immersed or nonimmersed.

plan a repeat in early 1999. But even as a pilot, the results exceeded our expectations. Clearly we have to do more work, but it seems that human subjects do respond appropriately to negative or positive audiences, even when these are entirely virtual. ■

Acknowledgments

The idea of applying VEs to social phobias was first suggested to us by Nathaniel Durlach, Senior Scientist at MIT’s Research Laboratory for Electronics, and Kalman Glantz a Boston-based psychiatrist. They made useful comments and suggestions throughout the study. This work is partially funded by the European ACTS (Advanced Communications Technologies) project, Coven (Collaborative Virtual Environments), and also the Digital-Virtual Center of Excellence project on Virtual Rehearsals for Actors.

References

1. D. Strickland et al., “Overcoming Phobias by Virtual Exposure,” *Comm. ACM*, Vol. 40, No. 8, 1997, pp. 34-39.
2. M.M. North, S.M. North, and J.R. Coble, “Virtual Reality Therapy: An Effective Treatment for the Fear of Public Speaking,” *Int’l J. of Virtual Reality*, Vol. 3, No. 2, 1998, pp. 2-6.
3. J.G. Tromp et al., “Small Group Behavior Experiments in the Coven Project,” *IEEE CG&A*, Vol. 18, No. 6, 1998, pp. 53-63.
4. F. Parke and K. Waters, *Computer Facial Animation*, A.K. Peters, Wellesley, Mass., 1998.

Contact Slater by e-mail at Dept. of Computer Science, University College London, m.slater@cs.ucl.ac.uk.

Contact department editors Rosenblum and Macedonia by e-mail at rosenblu@ait.nrl.navy.mil and Michael_Macedonia@stricom.army.mil.