David-Paul Pertaub
Mel Slater
Department of Computer Science
University College London
Gower Street, London WC1E 6BT

Chris Barker
Department of Psychology
University College London

An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience

Abstract

This paper describes an experiment to assess the anxiety responses of people giving 5 min. presentations to virtual audiences consisting of eight male avatars. There were three different types of audience behavior: an emotionally neutral audience that remained static throughout the talk, a positive audience that exhibited friendly and appreciative behavior towards the speaker, and a negative audience that exhibited hostile and bored expressions throughout the talk. A second factor was immersion: half of the forty subjects experienced the virtual seminar room through a head-tracked, head-mounted display and the remainder on a desktop system. Responses were measured using the standard Personal Report of Confidence as a Public Speaker (PRCS), which was elicited prior to the experiment and after each talk. Several other standard psychological measures such as SCL-90-R (for screening for psychological disorder), the SAD, and the FNE were also measured prior to the experiment. Other response variables included subjectively assessed somaticization and a subject self-rating scale on performance during the talk. The subjects gave the talk twice each to a different audience, but in the analysis only the results of the first talk are presented, thus making this a between-groups design. The results show that post-talk PRCS is significantly and positively correlated to PRCS measured prior to the experiment in the case only of the positive and static audiences. For the negative audience, prior PRCS was not a predictor of post-PRCS, which was higher than for the other two audiences and constant. The negative audience clearly provoked an anxiety response irrespective of the normal level of public speaking confidence of the subject. The somatic response also showed a higher level of anxiety for the negative audience than for the other two, but self-rating was generally higher only for the static audience, each of these results taking into account prior PRCS.

1 Introduction

Anxiety disorders are prevalent in the general population (Kessler et al., 1994) and can have a damaging influence on people's lives. The American Psychiatric Association manual DSM-IV defines a variety of anxiety disorders including panic disorder, obsessive-compulsive disorder, agoraphobia, specific phobia, and social phobia. This paper concentrates on fear of public speaking, a particular and common type of social phobia, and describes an experiment to

Presence, *Vol. 11, No. 1, February 2002,* 68–78

assess how speakers in front of a small virtual audience, in a seminar style setting, respond to different types of audience behavior.

Generally, people who suffer from social phobia have a strong fear of one or more social performance situations. They fear that they will act in a way that is humiliating or embarrassing and that others will judge them negatively. Although they recognize that this fear is irrational, they experience extreme discomfort and anxiety when in the feared situation and will seek to avoid the social encounter whenever possible. Fear of public speaking is a very common form of social phobia, with great social significance. People who fear speaking in public may find their career choices limited and avenues for promotion closed to them, resulting in considerable personal distress, frustration, and depression.

In the context of exposure therapy, patients exposed to a feared stimulus report that their anxiety attenuates with the passage of time, a process known as habituation (Emmelkamp, Bouman, & Scholing, 1995). Hence, the use of virtual environments to provide exposure in a controlled and inexpensive manner is a potentially useful application of this technology. A wide-ranging research effort has been investigating the possibility of using virtual reality exposure therapy to treat specific phobias in this way (Strickland, Hodges, North, & Weghorst, 1997).

The first studies involving VR exposure therapy were for acrophobia, a fear of heights (Rothbaum et al., 1995). Fear of flying has also been investigated (Rothbaum, Hodges, Watson, Kessler & Opdyke, 1996; Rothbaum, Hodges, Smith, & Lee, 2000), as has arachnophobia (Carlin, Hoffman, & Weghorst, 1997). Potential benefits were also found in the application of virtual treatment to agoraphobia (North, North, & Cobble, 1998) and claustrophobia (Botella et al., 1998). A recent validation study comparing the effectiveness of a low-budget virtual reality system and in vivo exposure for the treatment of fear of heights has found that both techniques were equally successful in reducing levels of anxiety and avoidance behavior in acrophobic individuals. Even more encouraging was the finding that, in both cases, the decline in fear of heights was sustained after six months (Emmelkamp et al., 2001).

To our knowledge, only one reported study applies virtual reality to the treatment of social phobia (North et al., 1998). Specially screened high-anxiety subjects were asked to give talks before a virtual audience of variable size that was programmed to laugh or ignore or encourage the speaker. After a five-week treatment session, there was evidence to suggest that virtual therapy was effective in reducing self-reported levels of anxiety, when contrasted with a control group who had been exposed to neutral stimulus virtual reality scenes. The virtual environment used in the public speaking trials involved a large auditorium and an audience of up to one hundred individuals. Most of the feedback from the audience was auditory, with limited visual appearance and actions of the individual virtual audience members. Clinical work is also being performed in this area, for example by Brenda Wiederhold (www.vrphobia.com/speaking.html), but, to our knowledge, results of clinical trials have not been published.

In this paper, we concentrate on small audiences and consider the prior question as to whether people's anxiety response is affected by the behavior of the virtual audience. The work of Reeves and Nass (1998) suggests that people do have a natural tendency to treat computers as social actors rather than mere tools or machinery, so it is likely that computer-generated avatars evoke a similar response. Also, in contrast to North et al. (1998), we concentrate on small audiences because it is likely that the smaller the size of the audience, the more the interaction approximates a conversational paradigm. In these situations, feedback from the listeners is crucial in facilitating the exchange of information and maintaining a sense of copresence. If an anxiety response in the speaker is generated in small seminar-like groups, the use of virtual reality is feasible in the limiting case of a single virtual listener and can be extended to a range of social performance situations.

The aim of the experiment was to investigate whether the type of virtual audience (hostile, friendly, or neutral) would affect the emotional response of the speaker. We describe the experiment in section 2 and details of the

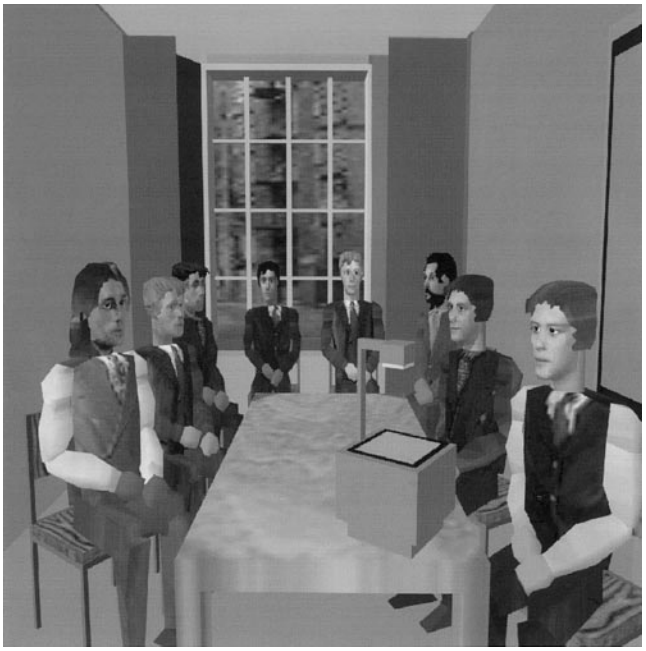

Figure 1. *The static audience.*

Figure 2. *The positive audience.*

variables used in section 3, with results in section 4. Conclusions are presented in section 5.

2 Experiment

2.1 Factoral Design

Altogether 43 subjects were recruited for the experiment by advertisement around the UCL campus and randomly allocated to three groups. Data were missing for three subjects, so the final number was forty. Each person gave two talks to a virtual audience. Group 1 gave a talk to an audience that was neutral in terms of emotional expression and entirely static throughout. These subjects then repeated their talk to the same static audience (figure 1). Group 2 first gave a talk to an audience that was friendly (figure 2) and then returned to give the talk to an audience that was hostile (figure 3). Group 3 first gave a talk to an audience that was hostile and then to an audience that was friendly. Between the two talks, the subjects completed questionnaires on their responses to the talk just given.

This could be treated as either a between-groups or a within-groups design. If only the results of the first talks are considered, it is a between-groups experiment. We allowed the possibility of within-groups as well in case there was insufficient power in the between-groups design. As it turned out, significant results were obtained from the between-groups design, and only that is considered in this paper. The type of audience was therefore one factor. A second factor was the level of immersion: the aim was for half of the subjects to experience the audience by use of a head-tracked head-mounted display (HMD) that placed the subject into a full stereo version of the seminar room. The other subjects viewed the audience on a standard workstation monitor (desktop).

The final distribution of subjects within the design is shown in table 1, including the breakdown by gender. Of the forty subjects, fifteen were undergraduates, seventeen postgraduate students, three were PhD students, two were postdoctoral, and the remaining three were other UCL staff. The average age was 28, with a similar age distribution in each of the cells of the experimental design. English was the first language of 23 of the subjects.

Figure 3. *The negative audience.*

Table 1. *Distribution of Subjects within the Design*

		Immersion		
Audience	Gender	Desktop	HMD	Total
Neutral	M	3	4	12
	F	3	2	
Positive	M	4	4	14
	F	2	4	
Negative	M	4	5	14
	F	3	2	
Total		19	21	40

2.2 Scenario

A virtual seminar room was populated with an audience of eight male avatars seated in semi-circular fashion facing the speaker, all dressed in suits as if attending a formal meeting. These avatars were continuously animated, displaying random autonomous behaviors such as twitches, blinks, and nods that were consciously designed to foster the illusion of a real-life presence (Vilhjalmsson, 1997). The room in which they were seated was a virtual counterpart of the real seminar room in which subjects completed their questionnaires.

Avatars in the positive and negative audiences responded to the speaker using one of three types of nonverbal communication: facial expressions, fairly static body postures held for more than a couple of seconds, and short animations (such as yawning, turning their heads, or walking out of the room). We simulated eye contact and listener gaze by enabling the avatars to look at the speaker and move their heads to follow the speaker around the room. Facial animation based on a linear muscle model (Parke & Waters, 1988) allowed the avatar to display combinations of six primary facial expressions with varying levels of intensity (Ekman, Friesen, & Ellsworth, 1972), together with yawns and sleeping faces.

A set of ten audience reactions was scripted for each of the two animated audience conditions. A selection of prerecorded verbal comments was also assembled for each condition, consisting of brief lexical items such as "I see," and "That's interesting" to more expressive evaluative phrases such as "That's absolute nonsense."

We used a selection of behaviors from the literature on nonverbal communication to design the negative and positive scenarios and included other actions that we assumed to project an unambiguous evaluative message (Argyle, 1988; Mehrabian, 1968; Bull, 1987; Veljaca & Rapee, 1998). In the negative scenario, avatars fell asleep, slouched in their chairs, slumped forward on the seminar table, oriented themselves away from the speaker, leaned backwards, put their feet on the table, and avoided establishing eye contact; one even got up and walked out during the talk. The appreciative audience was much more supportive, if less varied in the type and number of reactions. Avatars nodded encouragingly, smiled frequently, leaned forward, and oriented their bodies to face the speaker. They maintained eye contact with the speaker approximately ninety percent of the time. In one of the ten scripted responses, they clapped energetically, and, at the end of the talk, they gave the speaker a standing ovation.

In order to maximize the sense of copresence, we attempted to ensure some form of coordination between the responses of the audience and the content of the speech being given. An operator seated at a remote terminal unseen by the subjects used the distributive capabilities of the virtual environment to trigger the next reaction in the sequence at an appropriate moment. Only the timing, not the order, of the next animated audience response was at the discretion of the operator.

The audio responses could be triggered in any order and at any time, with the intention of providing a surrogate for the usual listener back-channel responses at appropriate points during the speech. The flexible timing of the animated responses and audio clips was deliberately incorporated into the scenario to foster a sense of interactivity and to avoid making subjects feel that the audience was responding at wholly inappropriate points in the talk. When designing the experiment, we could have triggered audio responses at fixed times throughout the duration of the talk or else allowed the operator to choose the timing. The advantage of the first strategy would have been to give each subject exactly the same auditory experience. However, this would have been at the cost, certainly for the positive audience, of giving auditory responses at potentially inappropriate moments, thereby making the positively responding audience potentially disruptive. The overall balance of auditory comments between the groups was the same on the average, although each subject did not receive precisely the same number of comments. This was not a drawback from the point of view of our goal: we were interested in whether subject responses were appropriate to the computer audience response, not the relative contribution of the audio and visual components of the scenario. So it would be enough to show that people who receive negative feedback become more anxious than people who receive pleasant and encouraging feedback. The maintenance of the positive audience as definitely positive required giving the operator control over the timing of feedback, and this was more important than achieving absolute equality of experience.

2.3 Method

Subjects were recruited from the staff and student body at the university. Subjects were asked to come to the Department of Computer Science on two occasions. On their first visit, the experiment was explained to them, they were shown the virtual reality equipment, and they were asked to complete several standard psychology questionnaires. Subjects completed the Personal Report of Confidence as a Speaker (PRCS) (Paul, 1966) and the Fear of Negative Evaluation (FNE) (Watson & Friend, 1969). These provided a measure of the subject's degree of apprehension when faced with the prospect of speaking in public. Subjects were also asked to complete the SCL-90-R (Derogatis, 1994), a standard symptom checklist that was used to screen for potentially confounding conditions. It has nine subscales representing the major psychological disorders (such as anxiety and depression). Subjects were told to prepare a 5 min. talk on a subject of their choice for presentation several days later. They were informed that they would be asked to give the talk twice to a virtual computer audience and that their presentations might be audio-taped for later analysis. Precise details of the workings of the system were not provided.

Upon return to the department, participants were asked to complete two more questionnaires before giving their talks. One of these questionnaires inquired about any anxiety they might have experienced while preparing for and anticipating the talk (using a modified version of the PRCS), and the other asked about their experience of social situations.

Participants using the HMD delivered their talk in a standing position alone in a darkened room set aside for the experiment, and the door was kept closed. Participants using the conventional computer terminal sat alone in front of a 21 in. monitor in the same darkened room, and again the door was kept closed. All participants were fitted with headphones and a microphone.

Subjects who gave their second talk to the negative audience were later asked to give a very short third talk, this time to the positive audience. This was done for ethical reasons so that they always concluded the experiment with a final positive experience. No data was collected for this talk.

When the subjects were in position in the presentation room, the concealed operator seated at a computer terminal in another room listened to the presentation and played a recorded invitation to the subject to begin the talk. At appropriate intervals, the operator generated the next in a sequence of predetermined virtual audience responses. In the case of the static audience, the operator simply monitored the duration of the talk. After the participants finished speaking, they were asked to complete a selection of questionnaires inquiring about their subjective experience of the talk.

At the end of the experiment, participants had the opportunity to discuss the experiment at greater length during a debriefing session. This allowed the researchers to gain insight into aspects of the experience of talking to a virtual public that are not easily captured in questionnaires. In this paper, we concentrate only on the statistical results. Results of the debriefing sessions are presented by Pertaub, Slater, and Barker (2001).

2.4 Screening

The SCL-90-R (Derogatis, 1994) was used to exclude subjects with significant levels of psychological disorder unrelated to fear of public speaking. We used criteria based on the population norms for this instrument:

- **For men:** All subscales (except psychoticism): exclude if any score is greater or equal to 1.5; psychoticism: exclude if this score is greater than 1.0.
- **For women:** All categories (excluding psychoticism): exclude if any score is greater or equal to 2.0; psychoticism: exclude if this score is greater than 1.5.

One person was excluded on this basis.

2.5 Materials

The virtual reality public speaking scenarios were designed using DIVE (Frecon & Stenius, 1998). The avatars themselves originated with DIVE, although significant customization was required, both to the geometry and fitting of texture maps. The avatar faces and accompanying muscle model originated with Parke and Waters (1998) but were texture mapped so that each face looked different. Similarly, clothing was texture mapped onto each avatar. The Robust Audio Tool (RAT) v3.023 (Hardman, Sasse, Handley, and Watson, 1995) was used to enable communication between the concealed operator and the speaker.

The experiments were conducted on a Silicon Graphics Onyx with twin 196 MHz R10000, infinite reality graphics and 192 MB of main memory. For the immersive sessions, the tracking system had two Polhemus FASTRAKs, one for the HMD and another for a five-button 3-D 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 deg. diagonal at 85% overlap. The screen resolution for the desktop system was 1280×1024 pixels. The concealed operator was seated at an SGI High Impact system with 200 MHz R4400 and 64 MB of main memory.

3 Variables

3.1 Response and Independent Variables

Several response variables were used. The first was a modified PRCS questionnaire (MPRCS). It was modified only by changing the tense so that the question specifically referred to the talk just presented, and by removing some questions that were not applicable in this situation. The MPRCS is based on sixteen statements, each with a yes/no response to be circled by the subject, and the count would be equal to 16 for a "maximal" level of reported fear of public speaking. This was administered after each of the two talks, and the modification was to make clear that the responses related only to the talk just given. The first five statements were, for example, as follows:

- I was in constant fear of forgetting my speech.
- At the conclusion of the speech I felt I'd had a pleasant experience.
- My thoughts became confused and jumbled when I spoke before the audience.
- I had no fear of facing the audience.

- Although I might have felt nervous before I began to speak, I soon forgot my fears and enjoyed the experience.

The second variable (somatic) was the reported somatic response indicating the degree of sweating (clammy palms, perspiration), discomfort in the stomach (butterflies), heart palpitations, tremors (shakiness), nerves/feeling of being scared, tightness in chest, tenseness, loss of balance, or nausea. Each item was ticked on a yes/no basis, and the maximum count of 9 would correspond to a maximally negative somatic response.

The third variable (self-rating) was the self-rating of the subject of the talk just presented, based on the question: "How do you rate your own performance in the talk you have just given? Assign to yourself a score out of 100. 0 = completely dissatisfied with your performance. 100 = completely satisfied."

The MPRCS and somatic variables were treated as the sum of binary responses, and thus analyzed with standard binomial logistic regression (McCullagh & Nelder, 1983). Binomial logistic regression is a very popular tool in the biological sciences, and it is appropriate to use whenever the response variable is the number of times a certain event occurs out of a fixed number of events. In this case, the event concerned was the choice in a binary response amongst a fixed set of questions. Logistic regression provides an analysis of how the count varies with other independent and explanatory variables. This tool has been used several times before in presence research, for example, by Slater, Steed, McCarthy, and Maringelli (1998). The self-rating variable was treated with normal regression.

The independent variables were the audience type (static, positive, negative), and the type of immersion (desktop display, HMD).

3.2 Explanatory Variables

Data were collected on a large number of additional variables: demographic—age, gender, ethnic origin, and first language; status—occupation (undergraduate, postgraduate, PhD student, postdoctoral, other); gaming—the degree of past computer game playing (never, sometimes, frequently).

A range of standard psychological questionnaires were administered prior to the experiment: PRCS, Fear of Negative Evaluation (FNE), and Social Anxiety and Distress (SAD) (Watson & Friend, 1969).

4 Results

4.1 Modified PRCS

The fundamental result can be seen informally by inspection of figures 4a through 4c. These show scatterplots of MPRCS (elicited after the experiment) against PRCS (measured prior to experiment) under each of the three types of audience. What is clear from the graphs is that there is a positive correlation between PRCS and MPRCS for the neutral and positive audiences (each of these is significantly different from zero at the 1% level), but the correlation for the negative audience group is not significantly different from zero. Moreover, the average MPRCS for this group is relatively high (7.7). In other words, for those who spoke to the neutral or positive audience, their response could be predicted from their normal level of anxiety about speaking in public. However, for those speaking to the negative audience, their response tended to be one of high anxiety irrespective of their normal level.

The variation in the scatterplots can be reduced by use of logistic regression analysis for the post-talk MPRCS. The best-fitting regression model found included the two main factors (audience type and immersion) and correlations between PRCS and MPRCS as described above. Gender also played a significant role in interaction with type of immersion. No term in the model could be removed without significantly worsening the overall fit at the 5% significance level. In the detailed results that follow, the "deviance" value (a goodness-of-fit statistic for logistic regression models) refers to the reduction in goodness of fit that would occur if the corresponding term were deleted from the model. This has an approximate *chi*-squared distribution with stated degrees of freedom. The details of the fitted model are as follows (other things being equal):

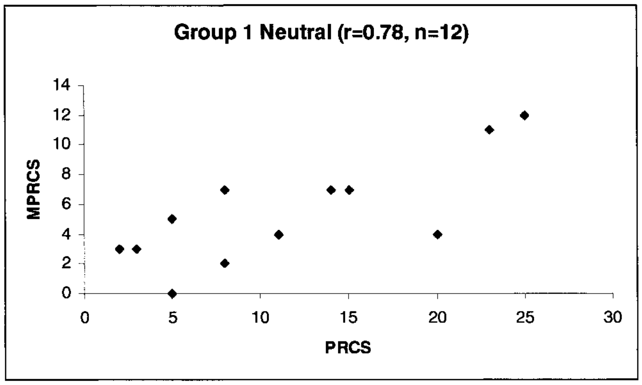

Figure 4a. *Plot of MPRCS against PRCS for the neutral audience.*

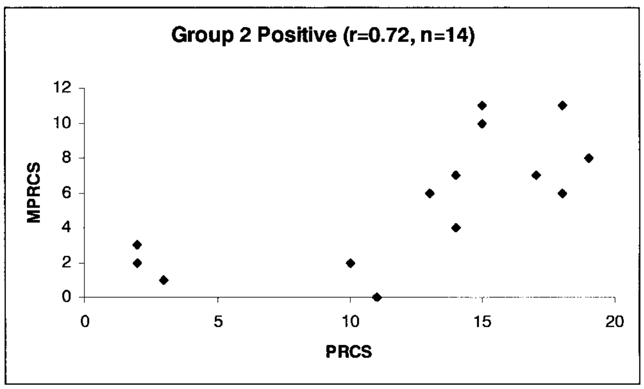

Figure 4b. *Plot of MPRCS against PRCS for the positive audience.*

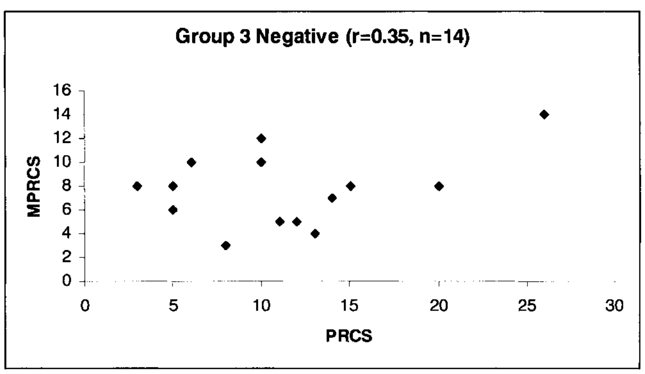

Figure 4c. *Plot of MPRCS against PRCS for the negative audience.*

1. The group that spoke to the negative audience has a significantly higher MPRCS, indicating greater anxiety, than the group that spoke to the static or positive audiences. (Change in deviance = 18.7 on 2 d.f.). There is no significant difference in MPRCS as between the static and positive audience.

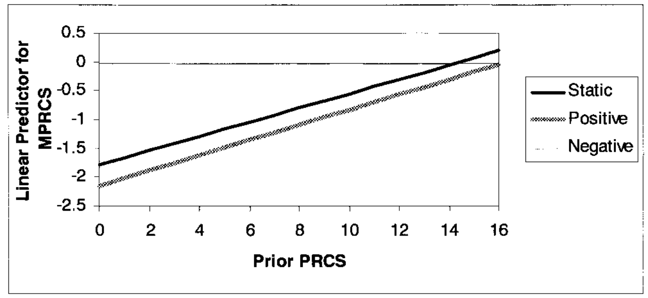

Figure 5. *Post-talk PRCS against prior PRCS for men using the desktop.*

2. MPRCS is lower for the group that used the HMD compared to the desktop group.
3. There is a significant interaction between gender and immersion. The higher level of MPRCS occurs for females using the HMD (change in deviance = 20.4 on 1 d.f.).
4. The slope of the regression line of MPRCS on prior PRCS is significantly different between the three audience types (change in deviance = 8.6 on 2 d.f.). For the static and positive audiences, the slope is positive (and not significantly different from one another). For the negative audience, the slope is not significantly different from zero. However, the intercept for the negative group is higher throughout the PRCS range than for the other two groups.

Result 4 is a very powerful and important result. What it says is that, for the people who spoke to the static or positive audiences, the best predictor of their post-talk PRCS is their pre-talk PRCS. However, for the people who spoke to the negative audience, their pre-talk PRCS cannot predict their post-talk PRCS, which is constant and high. This is illustrated in figure 5, which shows the fitted post-talk PRCS (linearized from the logistic regression) against the prior PRCS for each of the three audiences. The case shown is for the male subjects using the desktop, but each of the other cases shows the identical pattern. Note that there is no significant difference between the positive audience and static audience slopes and intercepts.

Table 2. *Means and Standard Errors for Self-Rating/100*

Audience	Mean response	S.E.
Neutral	67	6.4
Positive	48	5.9
Negative	41	5.9

4.2 Somatic

A logistic regression analysis was performed, and audience type and prior PRCS were significant (but nothing else). The reduction in deviance on deleting audience type from the regression was 15.2 on 2 d.f. The change in deviance for prior PRCS was 23.2 on 1 d.f.

The lowest level of somatic response corresponds to the static audience group, and by far the highest to the negative audience group. There is no significant difference between the positive and static group.

4.3 Self-Rating

A normal regression model (equivalent to ANOVA) was performed, and the results are similar to those for somatic. Audience type is significant, as is prior PRCS, but there is no difference between the slopes of PRCS amongst audience type. Generally, the higher the prior PRCS, the lower the self-rating as would be expected, but independently of audience type. The raw means amongst audience type alone are shown in table 2.

In this case there is no significant difference between the positive and negative audience results. It seems that self-rating is lower in the case that includes a "live" audience whether it is positive or negative.

5 Conclusions

This paper confirms the results of an earlier, much smaller pilot experiment reported by Slater, Pertaub, and Steed (1999). When people in a secluded empty room give a talk to a virtual audience, their response is affected by the behavior of that audience even though they know it to be virtual. The results of the current experiment—whether measured by post-talk PRCS or a subjective rating scale of somatic responses—show that a negative audience response provokes greater anxiety. The fundamental response variable of interest (the measurement of PRCS after the talk), which directly assesses anxiety in public speaking, shows a dramatic influence of the negative audience. Indeed, in this case, the fitted model shows that prior PRCS is uncorrelated with post-talk PRCS, which is constant and higher than for any other condition throughout its range.

There are some other interesting results. The impact of the greater immersion achieved with the use of a HMD was much higher for women than for men. To check the possibility that this was caused by extreme data points with undue influence on the regression lines, a standard leverage plot was constructed. This did identify the quite clear outlying point shown in figure 4c, but removal of the corresponding record did not change the result. Another possibility to consider is that there may be an effect of familiarity with virtual environments through computer game playing on the results; for example, it might be thought that women may be less likely to play computer games, and therefore more likely to be "impressed" by their experience. Each subject was ranked on a scale of 1 to 3 with respect to their level of game playing (1 being "never" and 3 being "often"). However, there was no significant difference in reported games playing between males and females. When game playing is brought into the logistic regression analysis for MPRCS, it is significant. There is an interaction effect with immersion, such that greater game playing is associated with an increase in MPRCS for those with the HMD but not for those using the desktop. (The change in deviance is 12.8 on 2 d.f.) It is very difficult to obtain an accurate estimate of the degree of computer game playing (how often, level of expertise, over what period, which types of games), so this should be taken only as an interesting finding for more serious consideration in future studies.

Subjects had the opportunity to talk about their experiences of the different types of audience in the post-experiment debriefing sessions. A discussion of the findings of these sessions can be found in Pertaub, Slater, and Barker (2001). Briefly, subjective reports confirmed that the negative audience was a strong anxiety-provoking experience and that it was capable of generating a range of very powerful emotions in the speaker. A further finding was that, despite the fact that real audiences are rarely so overtly hostile, the negative audience was frequently described as being the most realistic of the different scenarios. The static audience was least effective: viewing it on a desktop monitor, subjects commented that it was very like talking in front of a mirror.

We believe—from both the point of view of computer science and of potential therapeutic applications—that it is important to try to establish whether people do respond appropriately to virtual audiences. If this is the case (as our results strongly suggest), a virtual environment therapy can be successfully exploited. We have recently completed a further experiment on this issue with a greater number of subjects screened for high and low public speaking anxiety, comparing a lively but neutrally responding virtual audience to giving a talk in an empty virtual room. We were interested in contrasting the reactions of phobic and confident subjects and confirming that the presence of a virtual audience makes a difference to that reaction. The results show that social phobic subjects experience a much greater level of discomfort and anxiety on both objective (heart rate) and subjective measures when speaking to a group of virtual people than when speaking in an empty virtual room. Overall, this series of studies demonstrates that appropriate arousal can be provoked in people when speaking to audiences that they know to be entirely virtual, and this holds the promise that this type of approach might be successfully used in clinical interventions.

Acknowledgments

This research was funded under the European ACTS project Collaborative Virtual Environments (COVEN), and by a Wellcome Foundation project on Virtual Reality for Social Phobia. We would like to thank Prof. David Clark of the Institute of Psychiatry London for helpful comments and suggestions.

References

Argyle, M. (1988). *Bodily communication*. London: Methuen & Co Ltd.

Botella, C., Banos, R. M., Perpina, C., Villa, H., Alcaniz, M., & Rey, A. (1998). Virtual reality treatment of claustrophobia. *Behav. Research and Therapy, 36*(2), 239–246.

Bull, P. (1987). *Posture and gesture*. Oxford: Pergamon.

Carlin, A., Hoffman, H. G., & Weghorst, S. (1997). Virtual reality and tactile augmentation in the treatment of spider phobia: A case report. *Behav. Research and Therapy, 35,* 153–158.

Derogatis, L. R. (1994). *SCL-90-R adminsitration, scoring, and procedures manual (3rd edition)*. Minneapolis, MN: National Computer Systems.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1997). *Emotion in the human face: Guidelines for research and a review of findings.* New York: Pergamon Press.

Emmelkamp, P., Bouman, T., & Scholing, A. (1995). *Anxiety disorders. A practitioner's guide*. New York: Wiley & Sons.

Emmelkamp, P., Krijn, M., Hulsbosch, L., de Vries, S., Schuemie, M., & van der Mast, C. A. P. G. (2001). Virtual reality treatment versus exposure in vivo: A comparative evaluation in acrophobia. *Behav. Research and Therapy* [online] http://www.cg.its.tudelft.nl./~mertijn/vrphobia/vrinviv.pdf

Frecon, E., & Stenius, M. (1998). DIVE: A scaleable network architecture for distributed virtual environments. *Distributed Systems Engineering Journal, 5*(3), 91–100.

Hardman, V., Sasse, M-A., Handley, M., & Watson, A. (1995). Reliable audio for use over the internet. *Proc. INET'95, International Networking Conference*. 171–178.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., Wittchen, H.-U., & Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R Psychiatric Disorders in the United States. *Archives of General Psychiatry, 51,* 8–19.

Mehrabian, A. (1968). Relationship of attitude to seated posture, orientation and distance. *Journal of Personality and Social Psychology, 10,* 26–30.

McCullagh, P., & Nelder, J. A. (1983). *Generalized linear models.* London: Chapman and Hall.

North, M. M., North S. M., & Cobble, J. R. (1998). Virtual reality therapy: An effective treatment for the fear of public speaking. *International Journal of Virtual Reality, 3*(2), 2–6.

Parke, F., & Waters, K. (1998). Computer facial animation. A. K. Peters. Available at: www.crl.research.digital.com/publications/books/books.html.

Paul, G. (1966). *Insight vs. desensitization in psychotherapy.* Stanford: Stanford University Press.

Pertaub, D-P., Slater, M., & Barker, C. (2001). An experiment on fear of public speaking in virtual reality. In J. D. Westwood, H. M. Hoffman, G. T. Mogel & D. Stredney (Eds.) *Medicine meets virtual reality 2001.* (pp. 372–378). IOS Press.

Reeves, B., & Nass, C. (1998). *The media equation.* Stanford, CA: CSLI Publications; New York: Cambridge University Press.

Rothbaum, B. O., Hodges, L. F., Kooper, R., Opdyke, D., & Williford, J. (1995). Virtual reality graded exposure in the treatment of acrophobia: A case study. *Behav. Therapy, 26,* 547–554.

Rothbaum, B. O., Hodges, L. F., Kooper, R., Opdyke, D., Williford, J., & North, M. M. (1995). Effectiveness of computer-generated (virtual reality) graded exposure in the treatment of acrophobia. *American Journal of Psychiatry, 152,* 626–628.

Rothbaum, B. O., Hodges, L. F., Watson, B. A., Kessler, G. S., & Opdyke, D. (1996). Virtual reality exposure therapy in the treatment of fear of flying: A case report. *Behav. Research and Therapy, 34,* 477–481.

Rothbaum, B. O., Hodges, L., Smith, S., & Lee, J. H. (2000). A controlled study of virtual reality exposure therapy for the fear of flying. *Journal of Consulting and Clinical Psychology, 68* (6), 1020–1026.

Slater, M., Steed, A., McCarthy, J., & Maringelli, F. (1998). The influence of body movement on subjective presence in virtual environments. *Human Factors, 40*(3), 469–477.

Slater, M., Pertaub, D-P., & Steed, A. (1999). Public speaking in virtual reality: Facing an audience of avatars. *IEEE Computer Graphics and Applications, 19*(2), 6–9.

Strickland, D., Hodges, L., North, M., & Weghorst, S. (1997). Overcoming phobias by virtual exposure. *Communications of the ACM, 40*(8), 34–40.

Veljaca, K., & Rapee, R. (1998). Detection of negative and positive audience behaviours by socially anxious subjects. *Behav. Research and Therapy, 36,* 311–321.

Vilhjalmsson, H. H. (1997). *Autonomous communicative behaviors in avatars.* Unpublished master's thesis, MIT Media Lab. Available at: www.media.mit.edu/people/hannes/msthesis.

Watson, D., & Friend, R. (1969). Measurement of social-evaluative anxiety (FNE). *Journal of Consulting and Clinical Psychology, 33,* 448–457.