Social Presence Mediates Audience Behavior Effects on Social Stress in Virtual Public Speaking

Celia Kessassi, Mathieu Chollet, Cédric Dumas, Caroline Gl Cao

Social Presence Mediates Audience Behavior Effects on Social Stress in Virtual Public Speaking

Celia Kessassi
IMT Atlantique
LS2N, UMR CNRS 6004
F-44307 Nantes, France
0000-0001-6199-7578

Mathieu Chollet
School of Computing Science
University of Glasgow
Glasgow, UK
0000-0001-9858-6844

Cédric Dumas
IMT Atlantique
LS2N, UMR CNRS 6004
F-44307 Nantes, France
0000-0002-3779-0988

Caroline G. L. Cao
IMT Atlantique
Lab-STICC, UMR CNRS 6285
F-29238 Brest, France
0000-0003-1982-6489

***Abstract*—Public speaking is a near universally anxiety-inducing social situation. Applications recreating social interactions with autonomous agents in virtual reality have been proposed as tools to alleviate public speaking anxiety, recreating exposure therapy methods with virtual audiences. To be efficient, such applications rely on the precise induction of controlled amounts of social stress. We reviewed the literature that studied the effect of virtual audience behaviors on participants' stress levels, and found contradictory results that we attempt to explain in this article. We examine those studies and propose that social presence is one of the important factors mediating the effect of audience behavior on stress levels, and through social presence we can explain the different effects of audience behavior on stress observed in past studies. We conducted a study to test this theory, and expected that high social presence would lead to larger difference in stress induced by positive and negative audiences, as opposed to lower social presence which would attenuate the effect of audience behavior. We compared two display mediums, VR headsets to induce high social presence and wall projection to induce low social presence. We find an interaction effect of social presence and audience behavior on subjectively reported stress levels, but not on physiological measures.**

***Index Terms*—Social presence, co-presence, Virtual reality, public speaking, social stress, social evaluation**

I. Introduction

Public speaking is often reported as one of the most stress-inducing situations [9]. It is an activity that requires stress coping skills and practice. For some people, it can be debilitating to speak in front of audiences, especially in crucial situations, such as presenting at a conference, pitching a professional project, or performing a play. A large number of virtual training applications exist for participants to rehearse and prepare for such situations [1], [30]; however, little is known about the combination of factors (e.g., virtual agent behavior; graphical quality; training context, etc.) that make these applications effective. In particular, designing virtual simulations to train stress coping skills requires understanding the extent to which some factors are involved in stress. These applications must be able to induce the right amount of stress that users can learn to cope with. Many researchers have investigated different situations in virtual reality (VR) and have examined the extent to which they are stressful. However, a deeper look into such studies reveals some discrepancies in those examining the effect of audience behavior on participants' stress levels during public speaking in VR — specifically, whether virtual agents in the audience display positive or negative behaviors. [1], [26], measured stress through subjective and physiological measures and concluded that audience behavior moderates the levels of stress such that presenting to a negative audience leads to higher stress compared to presenting to a positive audience. However, even though [7] failed to replicate this effect, a very similar experimental design was used. We reviewed the three studies to analyze and compare the protocols and results, and conducted an experiment in an attempt to explain the contradictions. We propose that an important confounding factor, namely social presence, moderates the effect of positive and negative agent behavior on users and explains the differences across the three studies. Social presence refers to "the feeling of being there with a real person" [3], [24], and is higher when using immersive platforms compared to non-immersive ones, such as desktop platforms [19]. We believe immersive virtual platforms are more capable of delivering social cues than non-immersive platforms, therefore increasing their potential impact on social stress. In this paper, we present a study where we tested this hypothesis, comparing stress responses of participants with a VR headset versus a non-immersive wall projection condition.

This work was supported in part by the MISSIVES project funded by Agence Nationale de la Recherche (ANR) under reference ANR-20-CE33-0007, and the International Chair in Industry of the Future funded by the Region de Bretagne and FEDER, and the Univeristy of South Australia.

II. Related work

When designing a virtual application for public speaking training or exposure therapy, various conceptual choices come into the picture; for instance, the simulation's virtual agents (e.g., nod, smile, yawn, etc.) making up the virtual audience can affect participants' experiences, even mundane environmental changes may be potential factors affecting stress levels. These design choices can have implications on various cognitive mechanisms which in turn affect the effectiveness of the application. For example, [10], [16], [29] provided evidence that the level of social stress in VR is highly impacted by the sense of being observed. They compared the stress levels of participants who completed a speaking task in two different conditions, either they were being watched by virtual agents or humans, or there wasn't any sign of being observed in the environment. Only in the situations in which participants were being watched did the researchers observe increases in stress levels. In addition, several studies [1], [4], [27] highlighted the important role of behavioral realism in social interactions in virtual reality. For instance, inconsistent reactions from virtual humans may affect the user's level of engagement and stress (e.g., asking a question for which an answer was just given will disengage participants, and lower stress). Other studies [13], [17] proposed that real-life stereotypes apply to social interactions in virtual reality, including gender stereotypes and social status (e.g., expert vs. non-expert). Other factors, such as personal factors, including repeated past exposure to stressful virtual environments, or social phobia, are expected to influence stress levels, as indicated by [15], [32]. However, when it comes to the technical factors, [18], [33] suggest that a minimal level of graphical realism and immersion are enough for social stress to occur.

On the whole, these various studies paint a complex picture of inter-related factors involved in stress induction. In this study, we considered the valence (i.e., negative, positive) of virtual humans' behavior and its effect on social stress levels.

III. Analysis of 3 key studies

From the literature review, we identified three studies that specifically examined the effect of audience behavior on participants' stress levels [1], [7], [26]. In a seminal study on VR public speaking tasks by Pertaub et al. [26], participants were asked to speak on the subject of their choice to a virtual audience, and were told they would be audio-taped. During the talk, participants were either equipped with a head-tracked head-mounted-display (HMD) or looked at the audience displayed on a computer screen. The audience included eight male agents seated in a semicircular fashion. After the preparation period, participants gave their talk in one of three conditions, i.e., with agents displaying positive, negative, or static behavior. The positive audience showed interest, positive facial expressions such as smiling and nodding, and gave positive feedback (e.g., "That's interesting"). The negative audience displayed signs of inattention: falling asleep, orienting themselves away from the speaker, and leaning backwards. They displayed negative facial expressions and verbal feedback (e.g., "That's absolute nonsense."). The static audience did not include any animation or feedback. The authors used the Personal Report of Confidence as a Speaker (PRCS) [25] to measure participants' general confidence as speakers and a modified version (MPRCS) to measure subjective stress levels related to the talk. The results show that participants speaking to a negative audience reported higher MPRCS than participants speaking to a positive or static audience. The authors measured self-reported somatic responses such as the degree of sweating, shakiness, or heart palpitations. The results showed that the negative audience was associated with the highest self-reported somatic responses, and the lowest level was found in participants presenting to a static audience. For another measure, participants' self-assessments of presentation quality, no significant difference was found among the different conditions.

The second selected study was conducted by [1], the particularity of this study is the use of 360° recorded virtual audience. Participants were asked to improvise on three topics: studies or professional careers, a description of a place, and a description of a person. The virtual audience included ten male virtual agents seated in a semicircular fashion. Similar to [26], the positive audience displayed high attention with gaze directed at the speaker, nodding, leaning forward, and taking notes. The neutral audience was sitting upright, mostly looking down with occasional gaze at the speaker. The negative audience looked less interested, avoiding eye contact towards the speaker, leaning backwards with folded arms, using their phones and sometimes leaving the room. The results showed that self-reported anxiety was higher among participants who performed to a negative audience compared to those performing to a neutral one. Self-reported assessments did not differ significantly between the conditions. The authors also found Skin Conductance Levels (SCL) to be significantly higher in the negative condition than in the neutral condition. In contrast, Skin Conductance Responses (SCR) were not significant. Analysis of participants' Heart Rates (HR) yielded no significant differences. However, Heart Rate Variability (HRV) significantly increased in both positive and negative conditions compared with the control condition. In addition, order and time were associated with decreased SCL and HR levels and increased HRV. Voice analysis showed a decrease in intensity compared to the neutral audience and an increase in the percentage of silence compared to the neutral condition. Pitch, however, showed no significant difference.

In the study by [7], participants were asked to make four presentations. The audience was displayed on a large screen on a life-size scale. During each presentation, the audience's behavior followed a particular trajectory. For instance, the audience could start with high valence (positive behavior) and gradually change to neutral valence after 45 seconds. The authors expected the study participants' physiological measures to rise as the audience valence became more negative. However, there were no significant differences when analyzing the correlation between physiological measures (i.e., raw signals of Electrodermal Activity (EDA), HR, skin conductance response, and heart rate variability) and audience behavior. They also found no significant differences in self-assessments across conditions. Nevertheless, self-reported public speaking anxiety and HRV were significantly correlated, and it seems that prior anxiety level had an effect on self-assessment levels, with the higher the anxiety, the lower the self-assessment.

Table I is a comparison of the three studies where we can see the factors (i.e., topic of presentation, number of participants (N°), population with social phobia (SP)). From the table we can see that there are many similarities and a few factors which may explain the inconsistent results.

TABLE I: Comparison of the studies and their protocols

Study	***[26]***	***[1]***	***[7]***	***Current***
Design	Between	Within	Within	within
N°	40	36	28	36
SP	No	No	No	No
Topic	free choice	Studies, career, place, person	controversial topics	controversial topics
Audience behavior	Friendly, hostile, static - consistent	Positive, negative, neutral - consistent	High, low, neutral valence - varying	Positive, negative - consistent
Speech time	5 minutes	2 minutes	3 minutes	3 minutes
Behavior modalities	Facial expression, body movement, verbal feedback	Facial expression, body movement	Facial expression, body movement	Facial expression, body movement
Display system	VR, small screen	360° VR	Large screen	VR, wall projection
Measures	Subjective	Subjective, physiological, vocal	Subjective, physiological	Subjective, physiological, vocal
Results	effect	effect	no effect	effect in VR

IV. The hypotheses

These three studies try to answer the same conceptual question: does the behavior of virtual audiences affect speakers' level of stress during a public speaking simulation? They used very similar protocols, in which they asked participants to make a presentation to a virtual audience that exhibited a positive, negative, or neutral attitude. However, they do not provide the same answer; [26] and [1] provide evidence that there is an effect, while [7] reported finding no effect. When examining these experiments, we can see that differences exist in the experimental design (i.e., between or within), audience behavior design (i.e., 3D or 360 videos; consistent or varying behavior), in the type of measures (e.g., different questionnaires to measure stress), data analysis was conducted differently (mean comparison vs. correlation), and some displayed the audience in a VR headset, while others used computer screens. Various factors can explain these discrepancies, including low statistical power combined with the probability of having a type 1 or type 2 error. However, we argue that a confounding variable may be the primary reason of those differences, namely, social presence, which is itself mediated by the display system. We believe that participants speaking to a virtual audience in VR are likely to experience a higher sense of social presence. By amplifying the level of social presence, the behavior of the audience will influence the level of stress, i.e., making a presentation to a negative audience will lead to higher stress compared to a positive audience. In contrast, when participants make a presentation to a virtual audience displayed on a screen, the sense of social presence will be lower, therefore, the attitude and behavior of the audience won't affect the speaker as much.

To explore the interaction effect of social presence and audience behavior on stress levels, we conducted a similar study to the 3 we described above. This study explores the following hypotheses:

- **H1** - Performing public speaking to a virtual audience displayed in a virtual headset will lead to higher social presence compared to speaking to a virtual audience that is displayed in a screen projection.
- **H2** - Performing a public speaking task to a negative audience will lead to higher stress compared to public speaking to a positive virtual audience, but only when social presence is high, because there will be a higher influence of the audience's expressed attitude on the speaker.

V. Methods

A. Subjects

A call for participation was sent through the authors' university mailing list. The list comprises students, university staff members, and their entourage. Participants were invited to participate in an experiment that studied stress responses during a public-speaking task in virtual reality. 36 (F=24, M=12) French participants aged between 17 and 74 years replied to the call and took part in the experiment.

B. Task

The participants performed the task in an experimental room (8.70m x 5.45m). Half of the participants were randomly assigned to a Virtual Reality group (VR) where the audience was displayed in a head mounted display (HTC Vive Pro Eye). For the other condition, the audience was wall-projected (WP) in life-size (size:5.45m x 3.06m)(see Figure 1).

(a) WP

(b) HTC

Fig. 1: Experimental setup: virtual audience displayed using the screen projection, and an example of a participant wearing a head-mounted display.

A list of 12 public speaking presentation topics was constructed from social media polls. It includes controversial topics such as "Should the number of law enforcement officials be increased in France?" or "Do you support the demonstrations against the Covid19 vaccine pass?"

C. Material

1) Questionnaires: We administered several questionnaires at participant intake and after public speaking tasks. For all mentioned questionnaires, we either used a validated French version or created a translation for the experiment, depending on the available questionnaires.

a) Manipulation Check: Social presence (and co-presence) was measured using Nowak & Biocca's questionnaire [23] to check **H1**, i.e., whether social presence would be higher when participants interact with a virtual audience in VR as opposed as to a large life-size projection. Three sub-scales are included in the questionnaire: self-reported co-presence, perceived others' co-presence, and social presence [23] The perceived audience attitude (positive or negative valence and arousal [8]) was measured using a questionnaire. The questionnaire included 4 items (How [encouraging, bored, negative, interested] did the virtual audience appear to you?).

b) Primary measures: A visual analog scale (VAS) ranking from 0 to 100 was used to measure subjective stress before the experiment, just after the first presentation, and just after the second presentation. In addition, we used another subjective stress questionnaire, the STAI-Y1 [12] (which measures *state* anxiety), after each presentation.

c) Secondary measures: The STAI-Y2 [12] was used to measure *trait* anxiety level, which is a stable and long-term level of anxiety. We presume that individuals with higher trait anxiety will report higher stress levels after each presentation. We also collected additional questionnaires (Brief Fear of Negative Evaluation (BFNE), Brief Big Five (BFI-10) questionnaire, demographics), although we do not examine their associations with stress in this paper.

2) Physiological measures: Under stress, major physiological changes occur, including changes in the heart rate and skin sweat gland activity: as the stressed subject's arousal increases, their skin sweat glands activate and the released sweat increases skin conductance which can be measured electrically. We used a Galvanic Skin Response sensor (T-Sens GSR) [20] and a ECG sensor (T-Sens ECG), both of the sensors developed by TEA ERGO. For both modalities, we calculated a variety of measures, e.g. HRV, SCL, SCR and their associated features.

3) Behavioral measures: A webcam and a room microphone were used to capture posture, movements (and facial expressions in the screen projection case), prosody, and speech.

D. Virtual environment

To generate the virtual audience, we used the Cicero system [8] which generates virtual audiences expressing varying levels of arousal and valence through their nonverbal behavior. The audience comprised ten virtual agents, half female, half male, with varied age ranges, seated in a virtual conference room (see Figure 1). During the condition of positive behavior, the audience displayed high valence through a range of nonverbal behaviors, including smiling, nods, and leaning forward. In the negative behavior condition, the audience displayed low valence through frowning, head shakes, head-turning, and leaning backward. The system generates these behaviors dynamically according to the specified attitude to be expressed by the virtual audience, meaning that the exact sequence of nonverbal signals seen by a given participant is unique, but the model guarantees that it expresses the target attitudinal valence.

E. Protocol

After receiving approval from our University's ethical committee (CERNI), we conducted a study with a 2 × 2 factorial design in which the display medium (virtual headset vs. wall projection) and audience behavior (positive vs. negative) were varied. Participants were randomly assigned to a display group (virtual headset vs. wall projection) and had to complete a public speaking task twice, once with a positive audience and once with a negative audience. Upon arrival and after the informed consent form was signed, participants were equipped with ECG and EDA sensors, for which collected baseline measures over 5 minutes of rest. They were asked to fill in pre-test questionnaires, including the questionnaire for demographics, BFNE, BFI-10, STAI-Y2, and VAS for stress. They were then asked to choose a first topic from the list of 12 controversial topics, and they were told that they had five minutes to prepare for a presentation on that topic. After the preparation period, participants were invited to stand in front of the wall (approximately 2.5m away from the wall) and were equipped with the virtual headset if they were part of the virtual headset group. Once ready, the participants realised their presentations while having their performance recorded with the audio-visual and physiological sensors. The speech delivery phase lasted between 3 and 5 min, and participants were asked to stop early if their speech lasted longer than 5 min. During the presentation, the experimenter was isolated behind a privacy curtain, leaving participants to see only the virtual audience in the room. After each presentation, participants completed post-test questionnaires, including the questionnaire for measuring co-presence and social presence [23], the VAS and the STAI-Y1 for stress, and the questionnaire for perceived audience attitude. After completing the questionnaires, participants repeated the task, i.e., they were asked to choose a different topic and were given 5 minutes to prepare. If the audience's behavior was positive (resp. negative) in the first presentation, the audience's behavior in the second presentation was negative (resp. positive). Participants completed the same battery of questionnaires after the second presentation, and were compensated with a gift card of a value of 15€.

F. Data pre-processing

1) Scoring of the questionnaires: The STAI and BFNE scores was computed by summing the forward and reverse items, as indicated by [12]. Similarly, social presence scores (and co-presence) from [23] were computed by summing the forward and reverse items.

TABLE II: Timeline of the study - EXP1 = first presentation; EXP2 = second presentation; Co-P = co-presence; Soc-P = social presence; WP = Wall Projection group; VR = Virtual Reality

Group	**Pre-test**	**EXP1**	**Post-EXP1**	**EXP2**	**Post-EXP2**
VR/WP group assignment	BFI-10 BFNE STAI-Y2 Demographics VAS	Pos/Neg audience presentation	Co-P Soc-P Attitude VAS STAI-Y2	Neg/Pos audience presentation	Co-P Soc-P Attitude VAS STAI-Y2

2) EDA preprocessing and features extraction: Data from one participant were removed from the sample due to recording issues, resulting in a total of 35 participants with complete EDA data. The preprocessing started with a data-cleaning phase, following guidelines from [31]: All values near 0 (< 0.0001), values increasing by more than 20%, or decreasing by more than 10% in a 1 second window are removed and replaced by interpolation. In addition, all segments were plotted and visually inspected to check for noise. Afterward, a Butterworth low filter (0.5) was used to remove high-frequency artifacts, as indicated in [6]. The first three minutes of each recording were extracted (i.e., baseline, presentation 1, and presentation 2) and decomposed into phasic (Skin Conductance Reponse: SCR) and tonic (Skin Conductance Level: SCL) components, using Continuous Decomposition Analysis (CDA from Ledalab [2]). The following features were extracted: mean and standard deviations of the raw EDA, mean and standard deviation of the SCL, number of peaks, and higher peak.

3) ECG preprocessing and features extraction: All ECG data were plotted for visual inspection to check for data loss. Afterward, the data were cleaned and R peaks were detected using the Pan–Tompkins algorithm from the Neurokit package [21]. From this, the mean heart rate and heart rate variability (here, RMSSD: Root Mean Square of Successive Differences, a commonly used HRV feature [22]) were extracted.

4) Vocal features: Vocal features were extracted from the microphone recordings. Audio samples were analyzed using Parselmouth [14], a Python library based on Praat [5] to extract the fundamental frequency (pitch) and intensity. The pitch was extracted using a pitch floor of 75Hz and a pitch ceiling of 600Hz. The recordings of six participants were excluded due to recording issues.

VI. RESULTS

A. Manipulation Checks

We conducted a manipulation test to check whether the level of stress increased after the first and second presentations compared to the baseline. We conducted an ANOVA test to examine the main effect of condition (baseline, presentation 1, and presentation 2) on the level of self-reported stress (VAS). The test yielded a significant effect ($F(2, 70) = 24.192$, $p < 0.001$). A Kruskal-Wallis post-hoc test showed a significant difference($t(35) = -6.36$, $p < 0.001$) when comparing VAS at baseline with VAS at presentation 1 (95% CI $[17.11, 33.15]$); so did the comparison of VAS at baseline with VAS at presentation 2 ($t(35) = 4.93$, $p < 0.001$), $\Delta = 20.25$, 95% CI $[11.92, 28.57]$. A comparison of the VAS values between presentation 1 and presentation 2 showed no significant difference ($t(35) = -1.43$, $p = 0.16$, 95% CI $[-11.81, 2.03]$).

With our first hypothesis **H1**, we tested that a VR display system would lead to a higher social presence than the wall projection screen. To examine this effect, we conducted a t-test on social presence scores and compared them for the VR condition with those on the wall projection screen, and observed a significant difference on social presence ($t(69.86) = 2.48$, $p < 0.05$) and 95% CI $[0.93, 8.50]$. We also compared self co-presence, which showed no significant difference ($t(68.36) = -0.10$, $p = 0.91$) or 95% CI $[-5.30, 4.77]$, and perceived others' co-presence, which also yielded a non-significant result ($t(69.79) = -0.60$, $p = 0.54$) and 95% CI $[-4.64, 2.48]$.

We also checked physiological features to assess variations in stress levels at baseline and after the presentations. Only the number of peaks feature was significantly higher after both presentation 1 ($t(34) = 4.66$, $p < 0.001$) 95% CI $[9.49, 24.16]$) and presentation 2 ($t(34) = 5.56$, $p < 0.001$) 95% CI $[12.73, 27.37]$) compared to baseline. Among the extracted heart rate features, only mean heart rate reached significance: Mean heart rate showed a significant increase during presentation 1 ($t(35) = 4.12$, $p < 0.001$) 95% CI $[5.32, 15.66]$ and presentation 2 ($t(35) = 4.76$,$p < 0.001$) 95% CI $[4.98, 12.29]$ as compared with baseline (this could be due not just to stress but to a myriad of factors, including simply the higher effort demanded by the speaking task).

We also conducted a manipulation check on the perceived attitudes of the audience. An attitude questionnaire was used to check whether the audience's attitude in the negative condition (resp. positive) was perceived as negative (resp. positive). An ANOVA test for the interaction effect of behavior and display on the perceived attitude showed a significant main effect of audience behavior ($F(1, 34) = 59.677$, $p < 0.001$). When comparing the means of the positive and negative conditions, the test showed a significant p-value in the VR ($t(17) = -6.36$, $p < 0.001$), 95% CI $[-11.54, -5.79]$ and in the wall projection ($t(17) = -3.84$,$p = 0.002$) 95% CI $[-9.89 - 2.88]$. Results are shown in Figure 2.

The conclusions of our manipulation checks is that social presence was indeed higher in the VR condition as compared to the WP condition (H1 confirmed); further, the public speaking task was able to induce stress to participants, both subjective (VAS scores) and physiological (SCR number of peaks) stress. The negative and positive audience behaviors were also correctly identified by participants during the experiment.

B. Primary analysis

1) Self-reported stress: To examine self-reported stress levels, we conducted a factorial mixed ANOVA to compare the main effect of display medium (virtual headset vs. wall

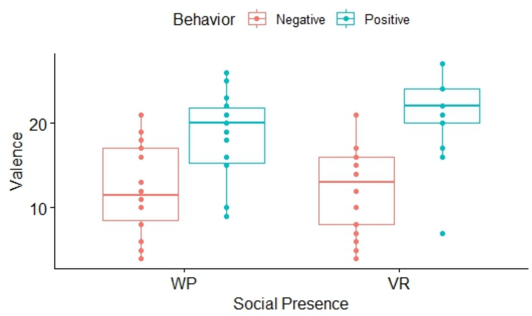

Fig. 2: Mean perceived valence depending on audience behavior.

projection) and the main effect of behavior (positive audience vs. negative audience) and their interaction on self-reported stress (VAS and STAI-Y1).

The mixed ANOVA revealed that there was no significant interaction between the display medium and audience behavior on VAS ($F(1, 34) = 0.950$, $p = 0.337$) and STAI-Y1 ($F(1, 34) = 0.019$,$p = 0.890$). However, there was a main effect of audience behavior on the VAS ($F(1, 34) = 5.898$, $p < 0.01$). Pairwise t-test showed the negative audience increased the VAS scores by 11 points an average (95% CI: [2.7–19.4]). No other significant effect was observed from the ANOVA analysis.

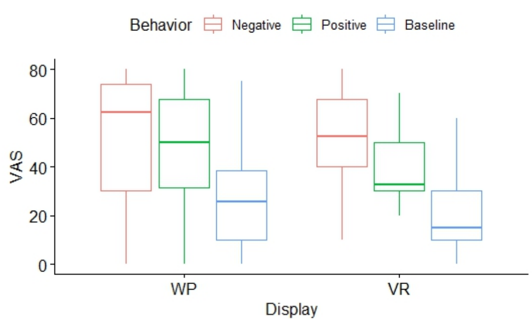

Fig. 3: Mean VAS by group and condition

2) EDA extracted features: Analysis of EDA features showed high variability across participants: by considering all the features (e.g., mean and standard deviation of SCL), a total of 11 outliers were identified. However, because this represents a high proportion of the sample, we decided to retain the data and continue the analysis. To examine the interaction between the display medium and audience behavior, we conducted an ANOVA on the extracted features. Only standard deviation (SD) of raw EDA values showed a significant main effect of behavior ($F(1, 33) = 4.725$, $p < 0.035$), however, the pairwise t-test did not show a significant p-value.

3) ECG extracted features: We conducted an ANOVA test on the extracted ECG features, including mean heart rate and heart rate variability (RMSSD). The results showed a near significant difference for the display medium on mean heart rate ($F(1, 33) = 3.755$, $p = 0.061$), no significant effect of behavior on mean heart rate ($F(1, 33) = 0.239$, $p = 0.628$), and no significant interaction effect between display and behavior ($F(1, 33) = 0.145$, $p = 0.706$). Regarding heart rate variability, a high number of outliers were identified. ANOVA analysis showed no significant main effect of display on heart rate variability ($F(1, 33) = 1.629$, $p = 0.211$), no significant main effect of behavior on heart rate variability ($F(1, 33) = 0.040$, $p = 0.843$), and no significant interaction term between behavior and display on heart rate variability ($F(1, 33) = 0.248$, $p = 0.622$).

4) Vocal features: We extracted the fundamental frequency (pitch) and intensity from the recordings. We conducted an ANOVA on pitch, and the results showed no main effect of display ($F(1, 27) = 0.181$, $p = 0.674$), no main effect of behavior ($F(1, 27) = 0.0007$, $p = 0.978$), and no effect of the interaction between display and behavior ($F(1, 27) = 0.048$, $p = 0.829$). Similarly, intensity was analyzed using an ANOVA test, and there was no main effect of display ($F(1, 27) = 0.775$, $p = 0.387$), no main effect of behavior ($F(1, 27) = 0.868$, $p = 0.360$), and no effect of the interaction between display and behavior ($F(1, 27) = 0.358$, $p = 0.555$).

C. Secondary analysis

In this section, we compare two groups: low and high social anxiety groups. The groups were formed by computing the STAI-Y1 score and separating high and low socially anxious group using a cutoff of 40, as suggested in [11]. The group size was 22 individuals in the high social anxiety group and 14 in the low social anxiety group. A t-test in the 95% confidence interval (CI) was used to compare the reported stress level of the high with the low SA group in the negative presentation showed no significant difference for the VAS scores ($t(28.954) = -1.5526$, $p = 0.1314$), while values of STAI-Y2 attained a significant p-value ($t(27.1) = -2.7156$, $p < 0.05$).

As part of testing hypothesis **H2**, we tested for an interaction effect between social presence and audience behavior. To do so, we separated participants into a group of high and a group of low social presence using the median as a cutoff. Then, we conducted an ANOVA test to test the interaction. The results showed that there was a significant main effect of audience behavior ($F(1, 28) = 5.367$, $p = 0.028$) and of the interaction term ($F(1, 28) = 4.929$, $p = 0.035$) over self-reported stress (VAS) (see figure 4). Pairwise t-test showed that there was a significant difference between VAS scores in the positive condition compared to the negative in the group with high social presence, while in the low social presence group no effect was observed. The t-test comparing VAS means between the negative and the positive condition in the high social presence group showed a p value marginally significant ($t(17) = 2.0517$, $p = 0.05$), 95%[-0.49, 35.49]. VAS scores in the negative condition was 17.5 points higher than the in the positive condition.

We realise another comparison on the two other sub-scales of [23] questionnaire: We conducted a t-test to compare the

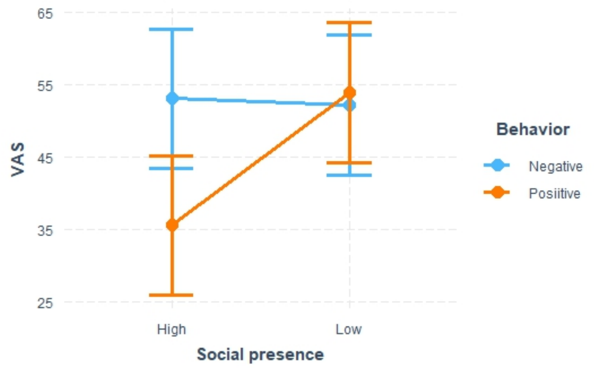

Fig. 4: VAS depending on social presence group (low: ”lowSo”, high: ”highSo”) and audience behavior (positive: +, negative: -)

scores in the positive condition with the negative condition and observed a significant difference ($t(35) = 5.1557$, $p < 0.001$), 95% CI $[7.16, 16.47]$: on average, self-reported co-presence in the positive condition was 11.81 points higher then in the negative condition.

VII. Discussion

We examined subjective, physiological and behavioral reactions to participants under a virtual public speaking situation. We first checked that our virtual public speaking task successfully triggered a stress response. Indeed, self-reported stress (VAS), number of SCR peaks and mean heart rate significantly increased during the presentations compared to baseline. Our main hypothesis (**H2**) proposed that the level of stress experienced by participants would be driven by audience behavior when social presence is high, while when social presence is low, the effect of audience behavior fades out. To induce high social presence we assigned participants to the VR group were the audience was displayed in head-mounted display, and to induce low social presence, we assigned participants to a wall projection group: checking this manipulation succeeded was the goal of (**H1**). Indeed, we observed a higher social presence in the VR condition compared to the WP condition, confirming **H1**, and we observed a near-significant interaction effect between social presence and audience behavior on subjective stress, as measures by VAS and shown in Figure 4. This partially confirms **H2**, although only in terms of subjective stress responses.

This result should however be interpreted carefully, and we want to highlight a number of caveats; first, the limited sample size of our study makes it important to analyse our results not as conclusive evidence but as findings to confirm in further experiments. Importantly, we must also stress that social presence itself was not controlled; instead it was the display medium of the audience. Also, we believe that the stimuli used in this experiment was relatively subtle and stronger stimuli could have led to higher stress standard deviations. Indeed, the audience stimuli included a variety of facial expressions, head movements and body postures, but stronger stimuli like verbal feedback may be more impactful. Finally, the nature of social presence itself and its relationship to socio-emotional responses is complex and may very well be bi-directional; [28] suggests a circular relationship between presence and emotions in virtual experiences. A secondary but interesting result of our study seems to confirm this, as the level of self-reported co-presence (i.e., how much participants rate themselves as engaged in the social interaction) was dependent on audience attitude; therefore it may well be that social presence and virtual audiences’ social behavior may be influencing each other in a complex manner, perhaps depending even on individual participant differences, such as immersive tendencies.

In the end, we believe this study confirms that it’s a complex challenge to provide a clear picture of the relationship between the factors of a virtual social interaction and experienced social stress, as shown by contradictory findings of past studies [1], [7], [26]. The multi-faceted psychological, physiological and behavioral effects of stress and the complex technological apparatus required for such studies means that most use different study design, measures, protocols, and that it seems nearly inextricable that such studies will differ on a variety of small to important issues.

VIII. Conclusion

In this study, we tested our hypothesis that social presence is an important factor driving the effect of virtual agents’ social behavior on participants’ subjective stress experiences, showing evidence of an interaction between social presence and audience behavior. However, this effect did not apply to other components of stress responses, including physiological measures of behavioral measures. Therefore, there is still much to research in order to understand well how the different design elements of a virtual social interaction application may impact user experiences. Research in this field may help us design better systems for exposure therapy or stress-coping skills training.

Ethical Impact Statement

The goal of our research is to enable applications which could lead to more efficient treatment approaches in the context of treatment social anxiety, social phobia and related disorders, as well as help in training public speaking skills. The approach we follow is generally based on exposure therapy, where patients are progressively exposed to the stimuli that poses them discomfort; this obviously poses an important ethical challenge as our experimental research methods require inducing social stress in participants - here to an extent not exceeding what could be experienced in daily situations - in order to elucidate the impact of different factors on stress induction itself. The well-being of participants being our principal priority, this research underwent ethical review by the university review board of Nantes University under grant n°27012022-1 which approved the study protocol, methods, and data management procedures prior to data collection commencing. To guarantee participant well-being, exclusion criteria were introduced to make sure that clinically anxious participants could not participate in the study; otherwise healthy volunteers that decided to proceed were informed of the risks of experiencing social stress and that they could withdraw at any time without penalty or justification; consent forms were collected and the experiment only proceeded when the experimenter was confident that participants understood this. The invitation letter informed participants that they would receive a 15€ gift card as a reward. At the end of the experiment participants received the compensation.

REFERENCES

[1] Miguel Barreda-Ángeles, Sara Aleix-Guillaume, and Alexandre Pereda-Baños. Users' psychophysiological, vocal, and self-reported responses to the apparent attitude of a virtual audience in stereoscopic 360°-video. Virtual Reality, 24(2):289–302, June 2020.

[2] Mathias Benedek and Christian Kaernbach. A continuous measure of phasic electrodermal activity. Journal of neuroscience methods, 190(1):80–91, 2010.

[3] Frank Biocca, Chad Harms, and Judee K Burgoon. Toward a more robust theory and measure of social presence: Review and suggested criteria. Presence: Teleoperators & virtual environments, 12(5):456–480, 2003.

[4] Jim Blascovich, Jack Loomis, Andrew C. Beall, Kimberly R. Swinth, Crystal L. Hoyt, and Jeremy N. Bailenson. Target article: Immersive virtual environment technology as a methodological tool for social psychology. Psychological Inquiry, 13(2):103–124, 2002.

[5] Paul Boersma. Praat, a system for doing phonetics by computer. Glot. Int., 5(9):341–345, 2001.

[6] Delphine Caruelle, Anders Gustafsson, Poja Shams, and Line Lervik-Olsen. The use of electrodermal activity (EDA) measurement to understand consumer emotions – a literature review and a call for action. Journal of Business Research, 104:146–160, November 2019.

[7] Mathieu Chollet, Talie Massachi, and Stefan Scherer. Racing heart and sweaty palms. In International Conference on Intelligent Virtual Agents, pages 83–86. Springer, 2017.

[8] Mathieu Chollet and Stefan Scherer. Perception of Virtual Audiences. IEEE Computer Graphics and Applications, 37(4):50–59, 2017.

[9] Erica Crome and Andrew Baillie. Mild to severe social fears: Ranking types of feared social situations using item response theory. Journal of Anxiety Disorders, 28(5):471–479, June 2014.

[10] Sally Dickerson, Peggy Zoccola, and Frank Zaldivar. Negative social evaluation, but not mere social presence, elicits cortisol responses to a laboratory stressor task. Health psychology : official journal of the Division of Health Psychology, American Psychological Association, 27:116–21, 02 2008.

[11] W. H. Emons, M. Habibović, and S. S. Pedersen. Prevalence of anxiety in patients with an implantable cardioverter defibrillator: measurement equivalence of the HADS-a and the STAI-s. Quality of Life Research, 28(11):3107–3116, June 2019.

[12] Janel Gauthier and Stéphane Bouchard. Adaptation canadienne-française de la forme révisée du state–trait anxiety inventory de spielberger. Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 25(4):559–578, October 1993.

[13] Alexandre Gomes de Siqueira, Heng Yao, Anokhi Bafna, Sarah Bloch-Elkouby, Jenelle Richards, Lauren B. Lloveras, Kathleen Feeney, Stephanie Morris, Erica D. Musser, Benjamin Lok, and Igor Galynker. Investigating the effects of virtual patients' nonsensical responses on users' facial expressions in mental health training scenarios. In Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology, VRST '21, New York, NY, USA, 2021. Association for Computing Machinery.

[14] Yannick Jadoul, Bill Thompson, and Bart de Boer. Introducing Parselmouth: A Python interface to Praat. Journal of Phonetics, 71:1–15, 2018.

[15] Peter Jönsson, Mattias Wallergård, Kai Österberg, Åse Marie Hansen, Gerd Johansson, and Björn Karlson. Cardiovascular and cortisol reactivity and habituation to a virtual reality version of the Trier Social Stress Test: A pilot study. Psychoneuroendocrinology, 35(9):1397–1403, October 2010.

[16] Oswald D. Kothgassner, Anna Felnhofer, Leon Beutl, Helmut Hlavacs, Mario Lehenbauer, and Birgit Stetina. A virtual training tool for giving talks. In Marc Herrlich, Rainer Malaka, and Maic Masuch, editors, Entertainment Computing - ICEC 2012, pages 53–66, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[17] Nicole Krämer, Sabrina Sobieraj, Dan Feng, Elisabeth Trubina, and Stacy Marsella. Being bullied in virtual environments: Experiences and reactions of male and female students to a male or female oppressor. Frontiers in Psychology, 9:253, 2018.

[18] Joung Huem Kwon, John Powell, and Alan Chalmers. How level of realism influences anxiety in virtual reality environments for a job interview. International Journal of Human-Computer Studies, 71(10):978–987, 2013.

[19] Christos Kyrlitsias and Despina Michael-Grigoriou. Social interaction with agents and avatars in immersive virtual environments: A survey. Frontiers in Virtual Reality, 2, 2022.

[20] Erika Lutin, Ryuga Hashimoto, Walter De Raedt, and Chris Van Hoof. Feature extraction for stress detection in electrodermal activity. In Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies. SCITEPRESS - Science and Technology Publications, 2021.

[21] Dominique Makowski, Tam Pham, Zen J. Lau, Jan C. Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and S. H. Annabel Chen. NeuroKit2: A python toolbox for neurophysiological signal processing. Behavior Research Methods, 53(4):1689–1696, feb 2021.

[22] Marek Malik, J. Thomas Bigger, A. John Camm, Robert E. Kleiger, Alberto Malliani, Arthur J. Moss, and Peter J. Schwartz. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. European Heart Journal, 17(3):354–381, 03 1996.

[23] Kristine L. Nowak and Frank Biocca. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. Presence: Teleoperators and Virtual Environments, 12(5):481–494, October 2003.

[24] Catherine S Oh, Jeremy N Bailenson, and Gregory F Welch. A systematic review of social presence: Definition, antecedents, and implications. Frontiers in Robotics and AI, 5:114, 2018.

[25] Gordon L. Paul. Insight versus desensitization in psychotherapy two years after termination. Journal of Consulting Psychology, 31(4):333–348, 1967. Place: US Publisher: American Psychological Association.

[26] David-Paul Pertaub, Mel Slater, and Chris Barker. An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience. Presence: Teleoperators and Virtual Environments, 11(1):68–78, February 2002.

[27] Sandra Poeschl and Nicola Doering. Virtual speech anxiety training — effects of simulation fidelity on user experience. In 2014 IEEE Virtual Reality (VR), pages 105–106, 2014.

[28] Giuseppe Riva, Fabrizia Mantovani, Claret Samantha Capideville, Alessandra Preziosa, Francesca Morganti, Daniela Villani, Andrea Gaggioli, Cristina Botella, and Mariano Alcañiz. Affective interactions using virtual reality: The link between presence and emotions. CyberPsychology & Behavior, 10(1):45–56, 2007. PMID: 17305448.

[29] Isabelle K. Sequeira, Addie S. Longmire, and Naomi J. McKay. Trier social stress test elevates blood pressure, heart rate, and anxiety, but a singing test or unsolvable anagrams only elevates heart rate, among healthy young adults. Psych, 3(2):171–183, 2021.

[30] Robin E. Sülter, Paul E. Ketelaar, and Wolf-Gero Lange. Speakapp-kids! virtual reality training to reduce fear of public speaking in children – a proof of concept. Computers Education, 178:104384, 2022.

[31] Sara Taylor, Natasha Jaques, Weixuan Chen, Szymon Fedor, Akane Sano, and Rosalind Picard. Automatic identification of artifacts in electrodermal activity data. In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, August 2015.

[32] Steven Taylor, Michael J Zvolensky, Brian J Cox, Brett Deacon, Richard G Heimberg, Deborah Roth Ledley, Jonathan S Abramowitz, Robert M Holaway, Bonifacio Sandin, Sherry H Stewart, et al. Robust dimensions of anxiety sensitivity: development and initial validation of the anxiety sensitivity index-3. Psychological assessment, 19(2):176, 2007.

[33] Sebastian Weiß, Nelly Klassen, and Wilko Heuten. Effects of image realism on the stress response in virtual reality. In Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology, pages 1–10, 2021.