Automatic Mechanisms for Measuring Subjective Unit of Discomfort

Dwi HARTANTO[a,1], Ni KANG[a], Willem-Paul BRINKMAN[a],
Isabel L. KAMPMANN[b], Nexhmedin MORINA[b], Paul G.M. EMMELKAMP[b],
Mark A. NEERINCX[a,c]

[a] *Delft University of Technology, Netherlands*
[b] *University of Amsterdam, Netherlands*
[c] *TNO Human Factors, Netherlands*

Abstract. Current practice in Virtual Reality Exposure Therapy (VRET) is that therapists ask patients about their anxiety level by means of the Subjective Unit of Discomfort (SUD) scale. With an aim of developing a home-based VRET system, this measurement ideally should be done using speech technology. In a VRET system for social phobia with scripted avatar-patient dialogues, the timing of asking patients to give their SUD score becomes relevant. This study examined three timing mechanisms: (1) dialogue dependent (i.e. naturally in the flow of the dialogue); (2) speech dependent (i.e. when both patient and avatar are silent); and (3) context independent (i.e. randomly). Results of an experiment with non-patients (n=24) showed a significant effect for the timing mechanisms on the perceived dialogue flow, user preference, reported presence and user dialog replies. Overall, dialogue dependent timing mechanism seems superior followed by the speech dependent and context independent timing mechanism.

Keywords. virtual reality exposure therapy, social phobia, anxiety level measurement, SUD score, speech recognition, speech detector.

1. Introduction

Social phobia is an anxiety disorder, where individual fear to do or say something in front of others that will be perceived as humiliating or embarrassing. The disorder is one of the most often occurring anxiety disorders, with reports that estimate this to affect 13.3% of the US population [1], 6.7% of European population [2], and 9.3% of Dutch population [3] during their live. The effect of this phobia on patients includes secondary depression and substance abuse (e.g. alcoholism, drug abuse), restricted socialization (e.g. professional, romantic and everyday informal social interaction), and poor employment and education performance [3]. Social phobia sufferers have a strong fear of social situations, such as talking in public, making a phone call, entering room with other people, ordering food in the restaurant, starting a simple conversation with strangers, etc.

[1] Corresponding Author. Email: d.hartanto@tudelft.nl

The gold standard for treatment of social phobia is exposure *in vivo*, where patients are gradually exposed to these social situations. One significant limitation of exposure *in vivo* is the difficulty for the therapist to get adequate and controlled social interactions (e.g. arrange the audiences, setup specific situation, etc.). Virtual Reality (VR) can overcome many of the shortcomings of *in vivo* exposure; in addition it provides a treatment that is more readily accepted by clients [4, 5].

During an exposure session, therapists normally ask patients about their anxiety level often using the Subjective Unit of Discomfort (SUD) score instrument [8]. SUD is a scale from zero ("no anxiety at all") to 10 ("the highest level of anxiety that you can imagine") measuring the subjective intensity or level of anxiety the individual is experiencing. With the aim of developing a home-based VRET system, where the system can be used for home treatment in which the patient can perform self-treatment without intensive therapist supervision, this measurement ideally should be done automatically using speech technology. However, with a VRET system for the treatment of social phobia, patients might be involved in a dialogue with an avatar in a virtual environment.

Since the SUDs measurement is done automatically with speech technology, a key question becomes the timing of asking for a SUD score as unexpected interruption might negatively affect patients' experience in a given situation, since poorly timed interruptions can adversely affect task performance [9, 10] and emotional state [11] of the users. To study the proper timing of asking participants to rate their SUD score in the dialogue-based virtual world, three proposed timing mechanisms were examined: (1) dialogue dependent (i.e. naturally in the flow of the dialogue, e.g. just before the start of a new avatar questions), (2) speech dependent (i.e. when both patient and avatar are silent), (3) context independent (i.e. randomly, but in this study when the patient is talking, testing a worst case interruption scenario).

2. Method

To study the proper timing of asking participants to rate their SUD score, 24 participants (11 females) were recruited in the study that was approved by the university ethics committee. The age of the participants ranged from 23 to 39 years (M = 30.3, SD = 4.7). All participants had a university background (current master and PhD students). Further, all participants had seen 3D stereoscopic images or movies, and none of them reported to have been exposed to virtual reality before. At the start of the experiment participants received a short introduction about the overall aim of the study and signed a consent form. After this, they completed the Personal Report of Confidence as a Public Speaker (PRCS) and the basic information questionnaire. Subsequently, speech recognizer was trained.

The main part of the experiment consisted of three sessions with the virtual audience of avatars, talking about three out of four different topics (chosen randomly from the following topics: Democracy, France, Dogs and Penguins [4]). To help them during the initial 3 minutes presentation about the topic, they were provided with a sheet containing some general pointers to talk about, which did not overlap with the question sets of the avatars. The presentation phase lasted 2 to 3 minutes, after which avatars started the question and answer phase which lasted around 1 to 2 minutes. All participants were exposed in a virtual environment using the Delft Remote Virtual

Reality Exposure Therapy (DRVRET) system [7] extended with implementation of the three different dialog timing interruption mechanisms for automatic SUD measuring.

The DRVRET system architecture was customized with a speech recognition and a speech detector engine interface. The speech recognition engine decodes and recognizes the speech from patients and then processes this further. In the current setup, DRVRET used Microsoft SAPI (Speech Application Programming Interface) 5.4 based on Windows 7 combined with the SPINX speech engine interface as its main speech engine. Speech detector functioned as a Voice Activity Detection, a technique used in DRVRET speech processing in which the presence or absence of human-avatar speech is detected. The software package Vizard was used for the visualization of the virtual room and avatars. Animations for avatars were done using 3D Studio Max using keyframe method. The hardware used was a Dell Precission T3400 with Intel quad core Q6700 @ 2.66 Ghz, 4 GB of RAM, with NVidia Geforce Quadro FX 4600 graphic card running on Windows 7 x64 bit and a Toshiba Satellite L300 laptop running on Windows 7 x32 bit. Participants sat behind a table equipped with microphone, facing a 3.5 by 2.5 meters virtual room projected with a screen resolution of 1280 x 1024 pixels at about two meters distance.

The experiment used a within-subject design and the order of three timing mechanism conditions was counterbalanced. In each session, participants were asked to complete the Igroup Presence Questionnaire (IPQ) [14], the Dialogue Experience Questionnaire (DEQ) [12] and the specially-designed questionnaire for this study to measure participants' experience after answering automatic SUD score questions: the SUD Score Experience Questionnaire (SEQ). During each exposure session participants were asked two to four times to rate their anxiety by giving SUD scores, depending on actual course of the dialogue between avatars and the participant.

3. Results

To study the effects of the timing mechanisms a series ANOVAs with repeated measures were conducted. A significant effect was found in the total SEQ score ($F(2,46) = 1065.24$; $p < 0.001$) and total DEQ score ($F(2,46) = 628.96$; $p < 0.001$). The total SEQ score (Table 1) suggests that participants rated the dialogue dependent timing mechanism as less interruptive than the speech dependent timing mechanism and the latter was again rated as less interruptive as the context independent timing mechanism.

The total DEQ score showed a similar pattern with regard to the dialogue experiences. Yet, an opposite pattern was found in the total IPQ score ($F(2,46) = 4.05$; $p = 0.024$). Participants rated presence highest for the context independent timing mechanism, while again the speech dependent in the middle and lowest for dialogue dependent timing mechanism. This might be a side effect of the phenomenon called Breaks In Presence [13] that participants might have experienced during the exposure, which occurs when they become aware of another reality. A possible explanation could be that the severity of the interruption made participants more aware of the break in presence switching from the virtual world to real world to answer SUD score questions, and back again to virtual world. Participants might have taken the intensity of break in presence as a sign of feeling present in the virtual world.

Table 1. Mean (SD) of measure for three timing mechanism.

Measure	Dialog Dependent	Speech Dependent	Context Independent
Total SEQ**	5.1 (0.6)	18.9 (2.2)	26.3 (2.0)
Total DEQ**	172.6 (3.3)	163.0 (4.5)	141.4 (4.0)
Total IPQ*	42.2 (3.5)	42.6 (3.5)	42.8 (3.5)

* $p < 0.05$; ** $p < 0.001$

During the experiment audio recording were made of the dialogues between avatars and participants. To understand the effect of the interruption for SUD score the analysis of the recording focused on how the participants continues talking with the avatars after they had given their SUD score. The participants' replies were coded with the following five labels: (1) detail answer, (2) normal answer (3) simple/short answer (4) "do not know" answer and (5) "lost in the dialog" answer. This resulted in five separate measures; each representing for a session the relative frequency of participants' replies that were coded with a specific label. With this coding scheme, three coders coded the audio recordings independently. Interobserver agreement was evaluated with Pearson's correlation analysis, showing (Table 2) acceptable agreement as all correlations were larger than 0.7 (all $p < 0.01$).

Table 2. Median (IQR) and Interobserver agreement of the relative frequency for five dialog replies in three conditions.

Measure	Dialog Dependent	Speech Dependent	Context Independent	Interobserver agreement (*r*)
Details answer	1.0 (1.7)	0.6 (1.2)	0.8 (1.3)	0.78 – 0.80
Normal answer	1.0 (1.9)	1.0 (1.9)	1.0 (1.7)	0.82 – 0.92
Simple/ short answer	0.7 (0.7)	1.0 (1.9)	1.0 (0.6)	0.80 – 0.89
"Don't know" answer*	0.0 (0.0)	0.0 (1.0)	0.0 (1.0)	0.98 – 1.00
"Lost in the dialog" **	0.0 (0.0)	0.0 (0.0)	0.0 (1.0)	1.00

* $p < 0.05$; ** $p < 0.001$

To test the overall effect of the timing mechanism on the participants' dialog replies, a series of Friedman test were conducted, using the five relative frequency measures as dependent measures (Table 2). No significant effect was found for the timing mechanism on participants' replies in the detail answer, normal answer and simple/short answer measures. However, a significant effect was found in the "don't know" answer ($\chi^2(2) = 6.05$, $p = 0.049$) and "lost in the dialog" answer ($\chi^2(2) = 22.00$, $p < 0.001$) measures. For more detailed analysis, paired comparisons with Wilcoxon Signed-Rank Tests were conducted on these two measures. In the "don't know" answer, only a significant differences was found between dialog dependent and speech dependent ($Z = -2.45$, $p = 0.014$) timing mechanism. Furthermore, in the "lost in the dialog" answer, significant differences were found between speech dependent and context independent ($Z = -3.03$, $p = 0.002$), and between dialog dependent and context independent ($Z = -3.03$, $p = 0.002$) timing mechanism.

4. Discussion and Conclusion

The results of the experiment seem to suggest that the automatic timing of asking participants to rate their SUD score could affect their subjective experience and their behaviour in a dialogue-based virtual world. Although potentially more development intensive, in-cooperating the moment of asking for SUD score into the flow of the

dialogue out performs other timing mechanism such as speech dependent and context independent timing mechanisms. Future research is needed to replicate these finding with social phobic patients. These findings can help developers to re-advance current VRET systems by implementing speech-recognition-based SUD score assessment, a feature especially relevant in future home-based VRET systems equipped with automatic feedback loop control system.

Acknowledgment

This research is supported by the Netherlands Organization for Scientific Research (NWO), grant number 655.010.207.

References

[1] R. C. Kessler, K. A. McGonagle, S. Y. Zhao, C. B. Nelson, M. Hughes, S. Eshleman, H.U. Wittchen, K.S Kendler, Lifetime and 12-month prevalence of DSM-III-R psychiatric-disorders in the United States - results from the national-comorbidity-survey.[Article]. *Archives of General Psychiatry*, **51**(1994), 8-19.

[2] L. Fehm, A. Pelissolo, T. Furmark, and H. U. Wittchen, Size and burden of social phobia in Europe, *Eur Neuropsychopharmacol,* **15**(2005), 453-62.

[3] R. de Graaf, M. ten Have, C. van Gool, S. van Dorsselaer, Prevalence of mental disorder and trends from 1996 to 2009. Results from the Nethelands mental health survey and incidence study-2. *Soc Psychiatry Epidemol*, 2011.

[4] A. Gracia Palacios, C. Botella, H. G. Hoffman, H. Villa, & S. Fabregat, Comparing the acceptance of VR exposure vs. in vivo exposure in a clinical sample. *Presented at Cybertherapy Conference*, San Diego, 2004.

[5] A. Gracia Palacios, H. G. Hoffman, S. See, A. Tsai, C. Botella, Redefining therapeutic success with virtual reality exposure therapy, *CyberPsychology and Behavor*, **4**(2001), 341-348.

[6] D. J. Katzelnick, K. A. Kobak, T. DeLeire, H. J. Henk, J. H. Greist, J. R. T. Davidson, , et al., Impact of generalized social anxiety disorder in managed care. [Article]. *American Journal of Psychiatry*, **158** (2003).

[7] W.P. Brinkman, D. Hartanto, N. Kang, D. de Vliegher, I.L. Kampmann, N. Morina, P.M.G. Emmelkamp, M.A. Neerincx, A virtual reality dialogue system for the treatment of social phobia, *CHI'12: CHI'12 extended abstracts on human factors in computing systems* (2012).

[8] J. Wolpe, The Practice of Behavior Therapy, ed 4. Pergamon Press: New York, 1990.

[9] B. P. Bailey, J. A. Konstan, J. V. Carlis, Measuring the effects of interruptions on task performance in the user interface, *Proc. SMC 2000* , IEEE, 757-762.

[10] M. Czerwinski, E. Cutrell, E. Horvitz, Instant messaging: Effects of relevance and time, In: S. Turner and P. Turner (Eds.), *People and Computers XIV: Proceedings of HCI 2000*, British Computer Society, **2**(200), 71-76.

[11] B. P Bailey, J. A. Konstan, J.V. Carlis, The effects of interruptions on task performance, annoyance, and anxiety in the user interface. *INTERACT '01*, 593-601.

[12] N. ter Heijden,, and W.P. Brinkman. Design and evaluation of a virtual reality exposure therapy system with automatic free speech interaction*, Journal of CyberTherapy and Rehabilitation*, **4**(2011), 41-55.

[13] M. Slater & A,J. Steed, A virtual presence counter*, Presence-Teleoperators and Virtual Environments*, **9**(2000), 413–434.

[14] T. Schubert, F. Friedmann, H. Regenbrecht, The experience of presence: factor analytic insight, *Presence: Teleoperators and Virtual Environments*, **10** (2001), 266 -281.