Preliminary version of Brinkman, W.-P., van der Mast, C., Sandino, G., Gunawan, L.T., and Emmelkamp, P. (2010). The therapist user interface of a virtual reality exposure therapy system in the treatment of fear of flying, *Interacting with Computers* , vol. 22, no. 4, pp. 299-310.

The therapist user interface of a virtual reality exposure therapy system in the treatment of fear of flying

Abstract

The use of Virtual Reality (VR) technology to support the treatment of patients with phobia, such as the fear of flying, is getting considerable research attention. Research mainly focuses on the patient experience and the effect of the treatment. In this paper, however, the focus is on the interaction therapists have with the system. Two studies are presented in which the therapist user interface is redesigned and evaluated. The first study was conducted in 2001 with the introduction of the system into the clinic. The original user interface design was compared with a redesign that was based on interviews with therapists. The results of a user study with five therapists and 11 students showed significant usability improvement. In 2008 a follow-up study was conducted on how therapists were now using the redesigned system. Using a direct observation approach six therapists were observed during a total of 14 sessions with patients. The analysis showed that: 93% of the exposures had similar patterns, therapists triggered 20 inappropriate sound recordings (e.g. the pilot giving height information while taking off), and more complex airplane simulation functions (e.g. roll control to make turns with the airplane) were only used by a therapist who was also a pilot. This resulted in a second redesign of the user interface, which allowed therapists to select flight scenarios (e.g. a flight with extra long taxiing, a flight with multiple taking off and landing sessions) instead of controlling the simulation manually. This new design was again evaluated with seven therapists. Again, results showed significant usability improvements. These findings led to five design guidelines with the main tenet in favour of a treatment-focused user interface (i.e. specific flying scenario) instead of a simulation-focused user interface (i.e. specific airplane controls).

Keywords: virtual reality; exposure therapy; fear of flying; mental health; design guidelines; user interface

1. Introduction

Fear of flying is not an uncommon phobia with reports that it affects 13.2% of the general population (Curtis, Magee, Eaton, Wittchen, & Kessler, 1998). Whether it is for business or personal reasons, flying has become a common mode of transportation in the industrial world. Avoidance of flying can therefore have both professional and social consequences. The diagnostic and statistical manual of mental disorders (DSM-IV-TR) (APA, 2000) classifies fear of flying as a situational type of specific phobia. The fear is marked and persistent and sufferers recognize that it is excessive and unreasonable. They try to avoid the situation or, as the manual states, endure it with intense anxiety or distress. A possible treatment is exposure therapy in vivo i.e. exposure in a real life situation. This therapy is considered the gold standard for treatment of phobias. With exposure in vivo patients are exposed to gradually more anxiety arousing situations for prolonged periods of time per session until anxiety dissipates and habituation occurs. Gradual exposure in vivo has been studied extensively (Emmelkamp, 2004). Exposure in vivo however has a number of limitations in that therapists do not always have full control over the real life situation (e.g. the weather condition) necessary to expose patients to a hierarchy of increasing difficulties; exposure can be difficult and expensive to arrange (e.g. flying in an airplane); and some patients are less willing to be exposed to situations they fear. Instead of exposing patients in vivo, exposing them in Virtual

Reality (VR) has successfully been put forward as an alternative. Recent meta-studies (Gregg & Tarrier, 2007; Parsons & Rizzo, 2008; Powers & Emmelkamp, 2008) show that exposure in VR is as effective as exposure in vivo. Furthermore, a specific meta-review on Virtual Reality Exposure Therapy (VRET) in the treatment of fear of flying (da Costa, Sardinha, & Nardi, 2008) concluded that this treatment is now an important reliable technique to be used in the treatment of this phobia. VRET effectiveness has also been studied for other phobias such as: claustrophobia, fear of driving, acrophobia (fear of heights), spider phobia, social phobia, panic disorder with agoraphobia, and Post Traumatic Stress Disorder (PTSD) (Krijn, Emmelkamp, Olafsson, & Biemond, 2004). Besides its effectiveness, research found patients more willing to expose themselves in VR than in vivo. For example in a survey (Garcia-Palacios, Botella, Hoffman, & Fabregat, 2007) among 150 patients, 76% chose in VR over in vivo exposure, but also importantly the refusal rate for in VR exposure (3%) was far lower than for in vivo exposure (27%).

As the Technology Acceptance Model (Davis, 1989) (TAM) points out, usefulness is a key factor for the acceptance of new technology. For more than a decade, therefore, researchers have studied the use of VR technology in the treatment of people with a phobia. Often the focus is solely on treatment effectiveness, and less on the efficiency by which a therapist administers a VR treatment session. However, as TAM also points out, the ease by which the technology can be used is also a key factor of system acceptance. Applying VR technology in a treatment puts an additional task load on therapists, as they have to interact both with the patient and with the VR system to create the exposure. The design of the therapist user interface seems therefore a key determinant of the therapist's task load. This paper presents work conducted over a nine year period which looks at several design iterations of a VRET therapist user interface used for treating people for their fear of flying. The paper starts with a discussion of the VRET system and the task of the therapist and the patient. The original therapist user interface is presented as a context for the first redesign and evaluation of the user interface that started in 2001. This is followed by the results of a series of field observations conducted in 2008. And this again is followed by the second redesign and evaluation of the therapist user interface. The paper concludes with presenting its main contribution put forward in five design guidelines, and reflects on lessons learned for the design of systems in the mental health domain.

2. Background

Collaboration that started in 1996 between Delft University of Technology and the department of Clinical Psychology at the University of Amsterdam resulted in a generic VRET system with VR worlds for the treatment of acrophobia, claustrophobia, and fear of flying. Both the patient and the therapist were identified as key interaction actors in the functional architecture of the Delft VRET system (van der Mast, 2006). Both users have their own user interface to interact with the VRET system. Patients are immersed in the virtual world wearing a Head Mounted Display (HMD) and their movements are tracked by the system. Speakers provide the audio element of the VR exposure. To enhance the feeling of presence even further patients sit in an airplane chair fitted with a bass amplifier to vibrate the chair during take off and landing. The therapist user interface consists of a screen on which the patient's view in the VR world is shown, and a screen with graphical widgets to control the system and record the session. Furthermore, the therapist user interface includes a keyboard, a mouse, and a joystick. This last one is to calibrate the patient's viewpoint. Besides the interaction with the system, patients and therapists also directly communicate with each other during a session.

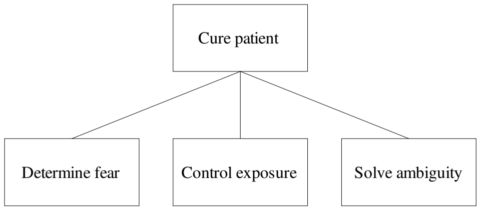

Figure 1: Therapist's goals, adapted from Schuemie (Schuemie, 2003).

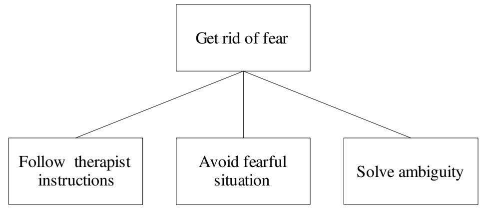

Figure 2: Patient's goals, adapted from Schuemie (Schuemie, 2003).

The set up of the architecture was initially based on a task analysis conducted on exposure in vivo therapy and later expanded to a VR setting (Schuemie, 2003). In the therapy, after an intake interview, patient and therapist first develop a hierarchy of fearful situations, and the goals that the patient likes to achieve. The exposure then starts with situations that the patient indicated as less fearful and continues until the fear has diminished, after which a more fearful situation from the hierarchy is selected for exposure. The therapist task (Figure 1) is to help patients to cope with their anxiety during the exposure session. To do this, therapists monitor the fear level by often asking patients to rate their anxiety by means of the Subjective Unit of Discomfort (SUD) scale (Wolpe, 1958), which ranged from 1 to 100 (or from 1 to 10). Based on the fear level, the therapist selects the appropriate exposure. The therapist also needs to respond to questions of the patient to solve ambiguity about the treatment and the phobia. For patients, their goal (Figure 2) is to get rid of their fear. To do this, they follow the therapist's instruction. However, they have a tendency to avoid fearful situations. Furthermore, to solve ambiguity patients also can ask clarification of the therapy during a session. For the therapy to work, the VR environment must elicit anxiety, which is not always achieved. For example, in her study (Krijn, et al., 2007) Krijn reported to have failed to accomplish this for 16 of the 50 participants. Presence is therefore seen as a key factor in the design of the patient user interface. Increased level of presence however will not always improve the effectiveness of the treatment. For example, although in the context of acrophobia, using a Computer Automatic Virtual Environment (CAVE) instead of a standard HMD resulted in a higher level of presence, however no improvement effect was found on the treatment (Krijn, Emmelkamp, Biemond, et al., 2004). Locomotion technique, e.g. walk-in-place, hand controlled viewing and gaze-directed steering, is another user interface factor influencing the patient's feeling of presence, whereby the more natural locomotion techniques contribute to higher levels of presence and fear (Schuemie, Abel, van der Mast, Krijn, & Emmelkamp, 2005). Still, patients being moved by the therapist can provoke more anxiety than when they are in control of their own movement (Schuemie, 2003). In the context of social phobia, dialogue techniques with virtual characters are also studied (Brinkman, van der Mast, & de Vliegher, 2008) as a presence factor, but also techniques for tracking the gaze direction of the patient to detect avoidance behaviour (Grillon, Riquier, Herbelin, & Thalmann, 2006).

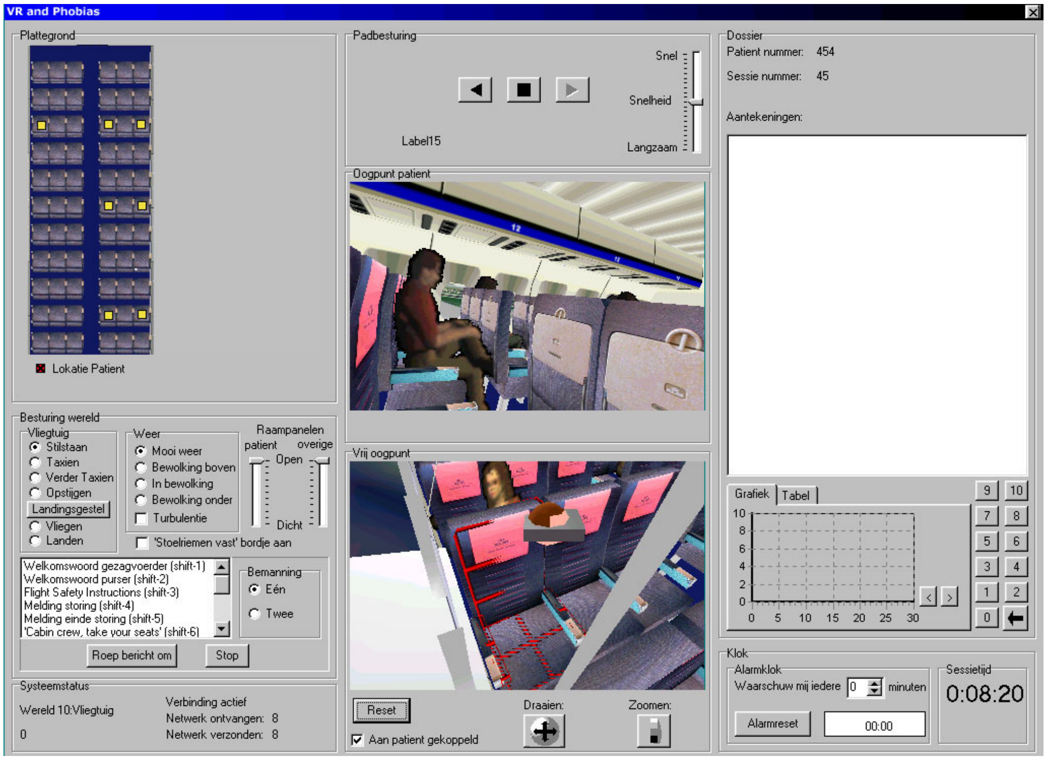

Figure 3: Original therapist user interface (in Dutch).

The therapist user interface is fundamentally different from the patient user interface as therapists and patients have different goals. Therapists have to cure the patient and to do this they do not have to be immersed into the virtual world or have a sense of presence. Instead, the therapist needs to control and keep an overview of the therapy. Figure 3 gives a screen shot of the original user interface for the therapist. Prominent elements in the therapist user interface are: the location panel with the plane seating map, control panel of the VR world (taxiing, taking off etc, weather type, windows open/closed, voice control pilot and purser), path control, patient view, free view, and patient information including notes, SUD scores, timer, and the session time. The design followed the guidelines set by Schuemie (Schuemie, 2003). He suggests that the therapist user interface also provides visual control feedback, e.g. graphical user interface with 2D and 3D control element, instead of only using keyboard and joystick commands that do not provide graphical information back to the therapist. Next, he also suggests that the user interface should allow recording and reviewing of SUD scores. Furthermore, he advises that the user interface should include graphical widgets to control the virtual environment and finally that the user interface should include a VR projection of real world objects, i.e. the patient, in the virtual world.

3. Initial Study

3.1. First redesign

By 2001 the system with the original therapist user interface was only used at two university locations. Before introducing the system to a clinic, a review of the system was undertaken, which included interviews with three therapists that had been using the system and a therapist of the clinic with no prior experience in using the system, but who was also a pilot. Based on these interviews a set of new requirements was formulated. (1) The system should allow flying to different destinations as patients need to be exposed more than once and multiple destinations would avoid patients getting used to one specific flight schedule. (2)

Preliminary version of Brinkman, W.-P., van der Mast, C., Sandino, G., Gunawan, L.T., and Emmelkamp, P. (2010). The therapist user interface of a virtual reality exposure therapy system in the treatment of fear of flying, *Interacting with Computers* , vol. 22, no. 4, pp. 299-310.

Next, the therapist should be able to slightly turn the aircraft so that patients have a better view from the airplane window to see the earth moving below them. This was expected to increase height perception, which is important as these patients often have acrophobia as well. (3) Therapists should also have more control over the flying conditions such as night or daytime flying, different weather conditions, but also control over the airplane such as cabin lights, and airplane sounds such as landing gear, flap wings, and turbulence. (4) Finally, a specific request was to make the therapist user interface easy to use.

The original therapist user interface was based on the design guidelines for a generic VRET system (Schuemie, 2003). Although consistency is an important usability principle (Nielsen, 2002), in this case the implementation seemed less appropriate. For example, the screen area for location control (upper left) is much larger than needed for the seating map of the plane, whereas the widgets to control the VR world were squeezed into small containers. Furthermore, the original user interface did not give controls that are more frequently used a more prominent place (e.g. SUD scores) than widgets that are less frequently used (e.g. note control). In the first redesign of the therapist user interface (Figure 4) this was addressed by designing a user interface specifically for fear of flying treatment.

In the original design, therapists could play any voice announcement at any moment during the session. To avoid situations where the therapist would trigger an inappropriate voice announcement (e.g. flight safety instructions while taking off), only flight stage appropriate voice control became selectable in the redesign. To remove the therapist's need to memorise the current stage the airplane was in, a flight view control was added to the interface (Figure 4, top right), indicating whether the aircraft was standing still, taxiing, taking off, flying, or landing. Furthermore, the therapist was given the ability to control the number of fellow (virtual) passengers (Figure 4, top second panel from left), roll the plane (Figure 4, middle bottom), set various weather conditions (Figure 4, lower left half), print a report of the session, and indicate that a session was started or ended including obtaining feedback on the duration of the session.

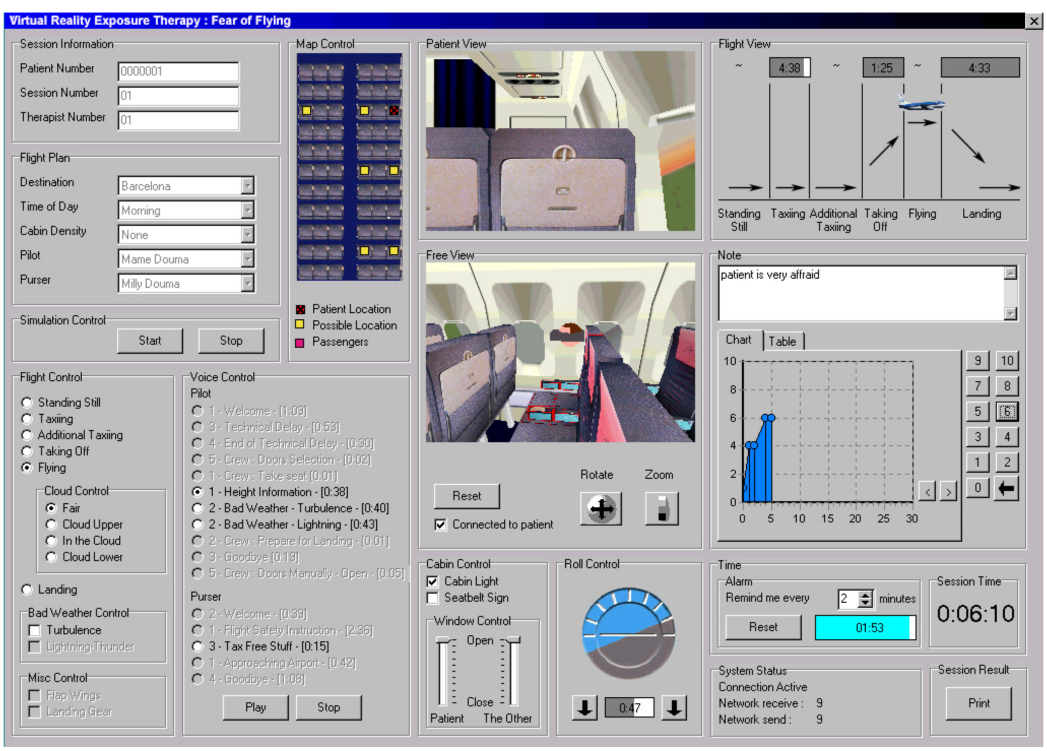

Figure 4: First redesign of therapist's user interface.

3.2. Evaluation of first redesign

In 2002 implementation of the first redesign was completed. An evaluation was conducted to examine the effect of the redesign and the new features on the usability of the system (Gunawan, van der Mast, Neerincx, Emmelkamp, & Krijn, 2004). Because of limited accessibility of the target group, only five actual therapists participated in the evaluation. However, the participant pool was extended with 11 master students who acted as therapists in the evaluation. They were following technology-oriented master degrees for example in computer science, electrical engineering, or applied physics. Beforehand the students received a small training session. Each of the 16 participants was teamed up with a student that acted as a patient, making the total of participants 32 in this evaluation. Each therapist participant conducted two treatment sessions: one with the original therapist user interface and one with the redesigned therapist user interface. To control for fatigue or learning effects, the order in which they used these two user interfaces was randomised. To avoid biasing participants by presenting one of the interfaces in a more favourable manner (Bentley, 2000), participants were not informed which system was the original and which the redesigned system. Each session took about twenty minutes, with a small break between the sessions. Participants received instructions to fill in patient and session details into the system and start gradually exposing the patient in the virtual world. After each session the 16 participants were asked to fill out a usability questionnaire (appendix A).

Of the 16 participants in the mixed student and real therapist pool, 10 preferred the redesigned system, five had no preference, and only one participant had a preference for the original user interface. This last participant considered the original design as less complicated due to the lower number of buttons that needed to be pressed to operate the system.

The results of the questionnaire showed internal consistency above an acceptable level of 0.7 (Loewenthal, 2001) with Cronbach's alpha of 0.91 for the original user interface and 0.93 for the redesigned user interface. Therefore, average rating of the questionnaire items was used as a usability index score. The factors (1) interface type and (2) being a student or a real therapist were analysed on their influence on the usability rating by conducting an ANOVA with repeated measures. User interface type was a within-subjects variable and participant type was a between-subjects variable. The results revealed that the participants in the mixed pool rated the usability of the redesign ($M = 1.55$, $SD = 0.76$) on average significantly higher ($F(1,14) = 16.32$, $p. = 0.001$) than the original user interface ($M = 0.84$, $SD = 0.90$) on scale from –3 to +3. The analysis found no main effect for participant type ($F(1,14) = 0.34$, $p. > 0.05$) nor an interaction effect between participant type and interface type ($F(1,14) = 0.40$, $p. > 0.05$). In other words, no difference was found between the rating of the students and the real therapists. This gives some justification of the used mixed participant pool. Still, this can also be caused by the small sample size. Therefore, a separate paired *t*-test on only the data from the actual therapists was conducted. It also revealed a significant ($t(4) = -4.04$, $p. = 0.016$) higher usability rating for the redesigned user interface ($M = 1.80$, $SD = 0.65$) than for the original user interface ($M = 0.93$, $SD = 0.52$).

The new features in the redesigned user interface were also evaluated by a set of separate usability and usefulness questions (appendix B). As Table 1 shows on a scale from –3 to +3, participants in the mixed pool gave the new features significantly higher scores than the middle scale value 0. This suggests that the new features were evaluated positively. When asked to list three things most or least liked in the system, the effects of lightning and turbulence was one of the favourite features, whereas less preferred features were the note taking feature and the timer alert to ask a SUD score that was not functioning very well in the redesigned system. In the debriefing phase the five real therapists were also given a separate questionnaire (appendix C) to evaluate the new features, including their estimation on how often they would use them. On a 5-point scale they rated the ease of use of the weather control, and the usefulness of the flight view control, the print function and the time feature significantly above 3, the middle value of the scale. Interesting was the five therapists' comments on how often they would use the roll control. One therapist indicated that he would use it often, while the other therapists did not. All the therapists agreed that the redesigned user interface was an improvement compared to the original user interface, and preferred it to treat patients with in the future.

Table 1: Usefulness and Usability results of the mixed student and therapist participant pool (n = 16) regarding additional features in the first redesign of the therapist user interface.

Feature	Usefulness	Ease of Use
Flight Plan Control		M^* = 1.94, SD = 1.34
Cabin Control		M^* = 1.88, SD = 1.03
Roll Control		M^* = 1.81, SD = 1.11
Flight View		M^* = 2.31, SD = 1.01
Print Function	M^* = 2.13, SD = 1.26	M^* = 1.88, SD = 1.50
Timer Feature	M^* = 1.50, SD = 1.32	
Simulation Control		M^* = 1.94, SD = 1.53

* *H_0: score = 0, p. < 0.001*

Table 2: Frequency of use, usefulness, ease of use of additional features in first redesign rated by real therapists (n = 5).

Feature	Frequency of use	Usefulness	Ease of use
Roll control	M = 2.2, SD = 1.3		M = 3.6, SD = 1.9
Weather control	M = 3.6, SD = 1.7		M^* = 4.8, SD = 0.4
Flight view	M = 4.0, SD = 1.7	M^* = 4.8, SD = 0.4	
Print Function		M^* = 4.4, SD = 0.9	
Timer Feature		M^* = 4.8, SD = 0.4	

* *H_0: score = 3, p. < 0.05*

The errors made by the participants in the mixed pool were also analysed. Errors in this context were defined as errors made by the 16 participants during the therapy session, and when assistance was needed. The data of each session was split into three phases: (1) loading the virtual world, (2) filling in therapy and patient information, and (3) exposing the patient. Analysis of only the second phase revealed that significantly (Wilcoxon signed ranks test, n = 16, Z = -2.45, $p.$ = 0.014) fewer errors were made. Only one participant made an error with the redesigned user interface, compared to seven participants with the original user interface. Whereas in the original user interface the information was entered in two separate windows, i.e. forms, in the redesigned user interface (Figure 4, upper left) therapists could enter this into the same form where they also control the simulation i.e. Figure 4 as a whole.

3.3. Discussion

The results suggest that the redesign of the therapist user interface was positively received and coincided with an apparent reduction of errors made during a session. Also the actual therapists were on average positive about the new features. Furthermore, they gave a number of suggestions for improvement. For example, they mentioned the idea of creating flight scenarios, which the therapist could simply select to run a session. Although they preferred this kind of autopilot, they wanted to maintain complete control over the session. Another suggestion was the introduction of cabin sound, e.g. people talking, baby crying etc. and an alert sound for alarm.

From a cognitive engineering point of view, it was interesting to see that no significant difference was found on the questionnaire response between the five real therapists and the 11 students. The number of therapists was low, which of course reduces the power of this statistical test. A test with more therapists would be needed to make any firm claim about this issue. However, access to this target group is difficult, and methods to substitute these types of participants in evaluations seem very welcome, likewise in other domains with similar accessibility restrictions (e.g. patients, physicians, military, but also astronauts).

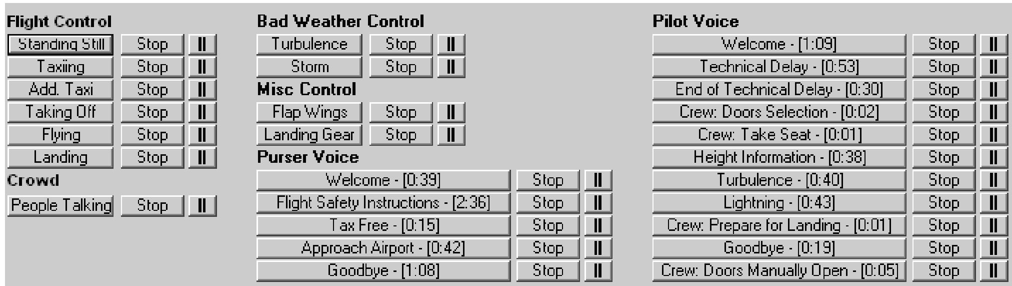

Figure 5: Extended control panel, placed below the first redesign of the therapist user interface.

In the years after this evaluation, the user interface, as requested by the therapists, was extended to include more cabin sounds such as flight safety instructions, and people talking in the background. As only the executable version of the system was available[1], a new sound panel (Figure 5) was placed below the user interface (Figure 4). To reduce redundant sound controls a grey panel was placed over the original voice control. The consequence of this technical update was that all sound recordings could be activated at any stage of the flight, something that was predicted to make the system more error prone.

4. Follow up Study

In 2008 a unique opportunity arrived. Because of some malfunctioning in the VRET system, some of the therapists in the clinic were less confident in using the system. They feared that the system would crash during a session and shatter patients' trust. After the system was repaired, these therapists were trained in using the system. In addition one of the researchers attended the initial therapy sessions as a technical assistant to build the therapists' confidence in using the system in case the system would collapse. Although the system did not collapse, it was soon realised that the researcher was in a unique position to observe how therapists actually used the system with real patients, and that this information could help to redesign the therapist user interface.

4.1. Field observations

In total 23 observations (Brinkman, Sandino, & van der Mast, 2009) were performed. Of these sessions, four did not include the VR flying simulation, but instead a VR world to treat patients for acrophobia. In two sessions the patients were exposed in a VR airport (and not the airplane) and in three other sessions the recorded data was incomplete. This reduced the data set to 14 sessions with 11 different patients and six different therapists.

Before their first VRET session, therapists first had an admission interview with the patient. In this interview patients were also trained in a number of exercises to relax the body during tension. The therapists explained that these exercises are important in order to give patients the feeling that they are in control of their own body. After the interview the therapist started with the actual VRET session, which included a short introduction to the VRET system. The therapist explained the functioning of the HMD, and how to adjust it. This procedure seems similar to the reports (Schuemie, 2003; Wiersma, Greeven, Berretty, Krijn, & Emmelkamp, 2008) of another Dutch clinic that also uses the same VRET system. After the introduction, the session started. The average length of a session was 25 minutes (*SD* = 3.1). In some cases the patients continue with an additional session. However, to avoid simulation-sickness, at least a break of a few minutes was always taken between two sessions.

[1] The vendor did no longer support the system in which the virtual worlds were modelled. So, in all stages exactly the same virtual world was used during this study.

A direct observation approach was applied and to ensure patient's privacy all recordings were done with pen and paper with no identifiable reference to the patient. As a treatment often includes multiple exposure sessions, observations included both patients new to VR exposure and patients that had already received a VR session. In the sessions, the observer recorded both the interactions between a therapist and the VRET system, and the direct communication between the therapist and the patient. An event sampling approach (Robson, 2002) was used to record the interaction with the system. The observer used a predefined coding scheme that referred to specific user interface controls, such as sound control, cabin control, and flight control. Besides the event sampling approach, a state sampling approach (Robson, 2002) was also applied by recording the phases of the flight, such as standing still (S), taxiing (T), additional taxiing (A), taking off (O), flying (F), flying fair (F1), flying below clouds (F2), flying in clouds (F3), flying above clouds (F4) and landing (L).

Table 1 shows the frequency of observed events. Taking the average for each therapist over his or her sessions, shows that the frequency in which the therapist spoke towards the patients (M =1.1, SD = 1.2) or asked SUD scores (M = 7.6, SD = 2.4) was significantly ($t(5)$ = 13.8, p. < 0.001) lower than the average frequency of the therapist interaction with the VRET system (M = 45, SD = 8.7). Several therapists explicitly also mentioned that the VRET system was demanding too much of their attention, as one put it 'the design had too many buttons'. Still, for the treatment a high level of interaction with the patient might not be desirable as it might reduce the level of presence in the VR world and therefore the exposure. Also interesting was that on average the time interval between a therapist's requests for two successive SUD scores was 3.6 minutes (SD = 1.1), significantly ($t(5)$ = -3.1, p. = 0.027) shorter than the often mentioned (Banos, et al., 2002; Wiersma, et al., 2008) 5 minute interval. However, some also report to use intervals of 2 minutes (Muhlberger, Herrmann, Wiedemann, Ellgring, & Pauli, 2001).

Table 3: Frequency of events during a session, session length, and order of phases.

Therapist	SUD asked	Patients comments	Therapist comments	Perform exercise	Voice control	Button pushed or selection made	Session length (min)	Order of phases
A	7	8	0	0	10	42	30	S T A O F1 F2 F3 F4 F1 L
B	5	3	3	2	8	37	24	S T O F1 F2 F3 F4 F1 F4 L
C	5	1	0	0	9	42	30	S T A O F1 F4 F3 F2 F1 L
	7	4	0	0	11	43	25	S T A O F1 F2 F3 F1 L
D	5	1	0	0	9	43	19	S T O F1 F2 F3 F1 L S
	12	3	2	2*	10	44	26	S T O F1 F2 F3 F4 F1 L
	8	0	1	1	9	40	20	S T O F1 F2 F3 F1 L
E	5	1	1*	0	9	32	21	S T O F1 F2 F3 F4 L
	5	1	1	1	6	35	17	S T O F1 F2 F3 F4 F1 L
	11	0	0	0	13	60	29	S T O F1 F2 F3 O F1 F2 F3 T O F1 F2 F3 L
F	13	2	4	1	15	67	27	S T A O F1 F2 F3 L T O O
	16	1	0	0	9	68	34	S T O F1 F2 F3 F4 F1 L T O F1 F2 F3 F4 F1 L
	9	4	0	0	11	54	23	S T O F1 F2 F3 F4 L
	10	5	4	4	15	59	19	S T O F1 F2 F3 F4 F1 L S

** At least, but could have been more; S - standing still, T - taxiing, A - additional taxiing, O - taking off, F - flying, F1 - flying fair, F2 - flying below clouds, F3 - flying in clouds, F4 - flying above clouds, and L - landing.*

Analysis of the changes between the flight phases showed that 93% of 14 sessions started with a similar pattern (start, taxiing (additional taxiing), taking off, flying, and landing). In

36% of the sessions the flight simulation was extended with another phase, for example taking off again. However, as Figure 6 shows that at the start of a session therapists followed a normal flying pattern.

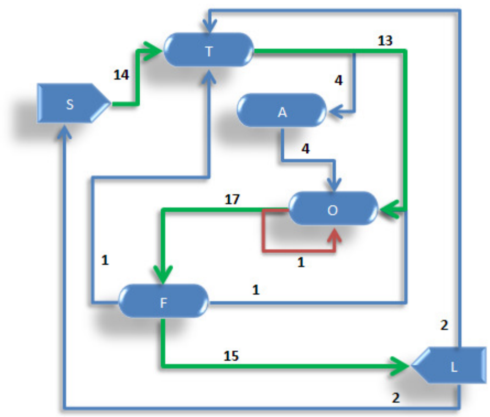

Figure 6: Sequence diagram of phases and the frequency of the phase transition observed (S - standing still, T - taxiing, A - additional taxiing, O - taking off, F - flying, and L - landing).

On average therapists played 9.9 (*SD* = 1.5) voice announcements from the pilot or purser. As mentioned before, the new sound panel (Figure 5) allowed therapists to play them whenever they liked. As expected, this had a negative effect. A total of 20 errors were observed where therapists initiated a sound recording inappropriate to the phase of the flight. This was especially the case for the taking off and the landing phase (e.g. the pilot giving height information while taking off or the pilot instructing the crew to open the door before a complete standstill). Several speculations can be made about the cause. First, therapists might have an inaccurate mental model of a flight or be unaware of the announcement's content (e.g. the welcome announcement where the pilot also mentions that the luggage is being loaded on board, played while taxiing). Or they have an accurate mental model, but they think that the airplane is in another phase of the flight, i.e. a mode error (Norman, 1981). Finally, there is also a more practical cause, therapists might have problems with fitting sound recording within a time slot of the phase, and therefore overshot a phase or start anticipating it and trigger sound recordings too early. Interesting was also the difference between therapists with and without experience in flying an airplane. The therapist with experience had a higher interaction frequency with the system (62 compared to the average 45), and used simulation functions other therapists never used (e.g. roll control to make turns with the airplane).

Some parts of the user interface were never or rarely used. For example, therapists did not enter any notes into the system. Instead they wrote down notes about the progress on the patient's paper form. As there was no printer attached to the system, notes recorded on the system would not be accessible after the session. However, therapists also explained they would try to avoid typing during a session, as the typing sound could distract the patient and therefore reduce the feeling of presence. Also therapists did not interact with the free view control, which allows the therapist to see a VR projection of the patient within the airplane.

4.2. Second redesign

In 2009, based on these observations a second redesign of the therapist user interface was made. This redesign was set out to address a number of issues. First the therapists should be provided with automatic fly scenarios. This would reduce therapist's task load during a session i.e. less interaction with the system. Next, with predefined scenarios, an actual pilot could put together a set of very accurate flight sessions, using also the more advanced options of the flight simulator such as roll control. Finally, appropriate sound recordings would be played in the right phase. A second issue involved note taking during the session. An improvement would need to avoid typing, but also again reduce the therapist's task load.

The approach of the second redesign was to obtain feedback early on in the design process. Therefore three paper prototypes were developed. All three prototypes included an automatic scenario control and an automatic note making control. With the latter control therapists could click on predefined notes such as 'the patient is tense', or 'relaxation exercise'. These notes were recorded in the log with a time stamp. The main difference between the prototypes was the graphical representation of the scenario control. Prototype 1 proposed an integrated solution in which both manual and automatic control elements were always visible. In the automatic control, therapists could see the separate phases of a scenario, including its current phase. In prototype 2 elements of the manual control such as sound recordings of the plane, the pilot and the purser were hidden in the automatic mode. The scenario was graphically presented as a list of blocks, representing different phases of the flight. Whereas prototype 2 focused on the entire scenario, prototype 3 only focused on the current and next phase. Therapists could see a timeline that illustrated the progress made in the phase. Flags were placed on the timeline to indicate when specific sound recordings would be played.

The low fidelity prototypes were used in a formative usability evaluation, where four real therapists interacted with the paper prototypes and whereby a researcher provided them with feedback of the system. This was either a piece of paper placed on the interface showing the new state of a control, or oral feedback on how the system would respond to their actions. By simulating the system it was hoped that therapists would get an idea of the interaction with the system, while clearly inviting them to speak openly at this early stage in the development process. The main observation of this evaluation was that the therapists wanted to automate all events in the scenario such as seatbelt sign, landing gear, or flap wings, and not only the pilot and purser announcements. They want to give a patient a flying experience without them needing to fly the airplane. They also were much in favour of having an overview of the entire scenario integrated within a timeline on which SUD score, and comment flags were placed including progress indicator of the flight scenario.

In addition to the evaluation with the therapists, the prototypes were also subjected to a heuristic evaluation (Nielsen & Molich, 1990). Six master students with background in Human-Computer Interaction (HCI) and two HCI lecturers reviewed the three prototypes. They were asked to list potential usability problems and prioritise them. The key usability issue identified concerned the location of controls on the screen, whereby related controls (e.g. because of the task sequence) should be located in close proximity on the screen. Especially, controls that could evoke more (or less) anxiety should be grouped together, giving therapists a clear overview of the options they have to control the patient's anxiety.

Based on the information obtained with the paper prototypes, a second redesign of the therapist user interface (Figure 7) was made. The screen was split up into four sections. (1) The upper left side of the screen was used to control pre and post session information such as patient information, session settings, but also the recorded log of the session. (2) The lower left side of the screen was used for general system information such as network status, but also SUD score reminders. (3) The upper right part of the screen was used to give an overview of the session plotted on a timeline, which includes recorded SUD score, comment flag set by the therapist, sound recordings, airplane events, and a progress indicator. (4) The lower right side of the screen was used for controls therapists can use during a session such as moving to the next phase of the flight, inputting SUD scores, setting comment flags, controlling manual simulation events such as cabin light, or window shutters, and controlling specific elements to create a more fearful exposure such as bad weather, turbulence, and turning the aircraft, all placed together in a single anxiety panel. The design also introduced the use of tabs to switch between controls often and less often used, for example between the timeline and the control to create and select a scenario, but also between automatic and manual control. Although therapists could switch between the manual control tab and automatic control tab, they needed to press a separate button to actually activate a new mode. This was done to reduce the risk of accidental activation of the manual mode in the middle of a session. Still in the manual mode, the user interface only allowed therapists to play flying phase appropriate sound recordings. Therapists could select a specific phase of the flight and choose to trigger sound recordings manually, or let the computer do this.

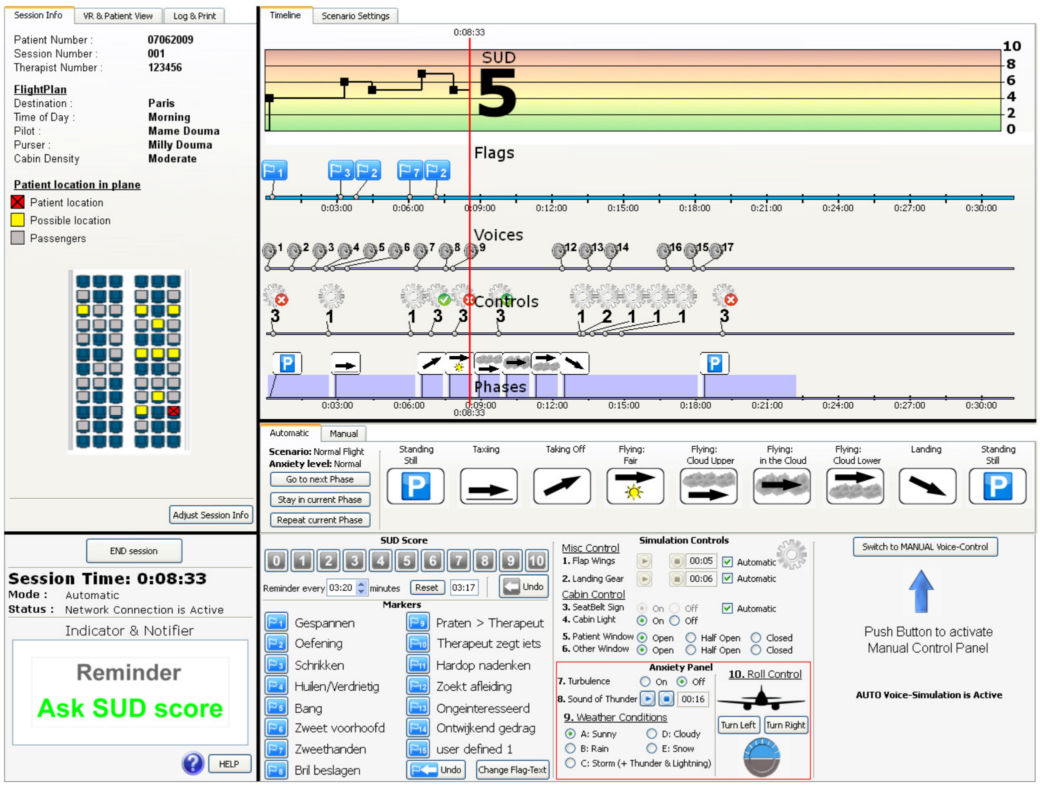

Figure 7: Second redesigned therapist user interface.

4.3. Evaluation of second redesign

To evaluate the usability of the second redesign a so-called Wizard of Oz set up was used, whereby actual therapists used the new user interface, but where a researcher manipulated the actual VRET system behind a screen. The researcher also played the role of the patient, giving therapists oral feedback on how the patient felt including SUD score information. The response of the patient was written out before the experiment to have some degree of consistency in the patient's reaction between the sessions. All seven therapists that participated in the experiment had prior experience in using the VRET system. Six worked in a clinic and one at a university. The therapists completed two sessions with the VRET system, one in manual mode and one in automatic mode. To control for fatigue the order of the two sessions was counterbalanced. After each session, therapists completed a component-based usability questionnaire (Brinkman, Haakma, & Bouwhuis, 2009) concerning the specific manual or automatic control components, and at the end of the experiment therapists completed a component-based usability questionnaire regarding the remaining components, as well as an adapted version of usability questionnaire used in the initial evaluation (appendix A) in 2002. With the latter questionnaire, therapists also evaluated the user interface of the current VRET system (1st redesign, Figure 4, extended with the sound control panel, Figure 5).

Table 4: Reliability of the usability questionnaire items used in the evaluation of second redesign and results *t*-test with test value = 5.29 (*n* = 7).

Item	Cronbach's α	*M*	$t(6)$	*p*
Automatic Phase control (automatic mode)	0.84	6.6	8.86	.000
SUD score, comment flags, simulation control, and anxiety panel (automatic mode)	0.93	6.2	3.17	.019
Manual phase control (manual mode)	0.97	6.4	5.91	.001
SUD score, comment flags, simulation control, anxiety panel, and voice control (manual mode)	0.99	5.0	-0.42	.687
Session information control	0.97	6.8	10.95	.000
General system information control	1.00	6.9	13.42	.000
Timeline control	0.97	6.1	2.54	.044
Therapist's user interface usability questionnaire (first redesign with sound extension)	0.93	0.4		
Therapist's user interface usability questionnaire (second redesign)	0.92	1.8		

Table 4 shows that the questionnaire elements obtained acceptable level of reliability with Cronbach's alpha above the threshold level of 0.7 (Loewenthal, 2001) ranging from 0.84 to 1.00. This made it possible to continue the analysis focusing only on the aggregated measures. The ratings of user interface components established with component-based usability questionnaire were compared with the norm value of 5.29 (Brinkman, Haakma, et al., 2009). Ratings above this value suggest that the component's usability is more similar to the usability of the easy to use components in the norm set, and likewise components with rating below this value are considered more similar in usability to the difficult to use components in the norm set. As Table 4 shows, the therapists rated all components significantly above this threshold, with the exception of the component that also included the manual voice control. As the rating for that component was neither found to be significantly below the threshold, it should not be interpreted as an unusable component. Although setting comment flags was new to them, therapists mentioned that they liked this functionality. They also liked the automatic scenarios, as they thought that this would reduce their task load. One therapist wrote down that she liked the automatic scenarios and added 'I am not a pilot, so I wouldn't forget anything'. This seems to confirm that therapists no longer required having a detailed mental model of a flight. Furthermore, the therapists also scored the usability of the second redesign significantly higher ($t(6) = 5.34$, $p. = 0.002$) than their current user interface.

As the same questionnaire data was collected in the initial usability evaluation it was possible to compare the usability rating of all the user interface designs by actual therapists. Figure 8 shows the average rating on questions included in both questionnaires. Interesting is the drop in the usability rating between the first redesign in 2001 and the rating of adjusted first redesign in 2009. Although this might be caused by the implementation of the sound panel, other factors might also play a role, such as (1) in 2009 therapists all had extensive experience in using the VRET system whereas in 2002 this was not the case; (2) both evaluations had a within-subject design, therapists might have been compelled to give relative and more extreme ratings; and (3) therapists might have given socially desirable ratings thereby seeing a new interface as more favourable (Bentley, 2000). Therefore, instead of only considering the subjective data, examining the therapists' interaction with the system in the manual and automatic mode gave an indication of possible reduction of the task load. The therapist made significantly fewer ($t(6) = -7.72$, $p. < 0.001$) mouse clicks when treating a patient in the automatic mode ($M = 26$, $SD = 10.5$) than treating a patient in the manual mode ($M = 64$, $SD = 11.1$). The amount of user interface actions in the automatic mode was also significantly lower ($t(6) = -4.67$, $p. = 0.003$) than the average of 45 interactions observed in the 14 field observations with the adjusted first redesign of the user interface. Noteworthy was also the number of pilot and purser voice recordings played ($M = 14$, $SD = 2.8$) in the manual mode, which was significantly ($t(6) = 3.92$, $p. = 0.008$) higher than the 9.9 average observed in the field earlier. This could be simply been an artefact of the experimental setup. However, therapists might have been more willing to select them, as they probably knew they were appropriate for the current phase of the flight.

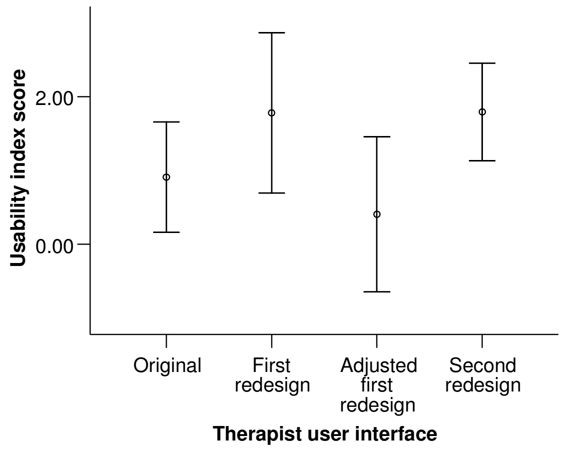

Figure 8: Actual therapists' mean usability rating of various user interface design ranging from –3 (low usability) to +3 (high usability), including a 95% confidence interval.

When asked which user interface they would like to use in the future all seven therapists preferred the second redesigned interface. Therapists also liked the idea that they could personalise the scenario. Using scenarios in a research environment was also seen as a way to obtain more consistency between the treatments. This would reduce individual variability within a test condition and makes it easier to detect differences between test conditions. Or in other words, might potentially increase the statistical power of an experiment.

5. Discussion and conclusion

As with any empirical evaluation, the evaluation also had a number of limitations. For example, not all evaluation sessions included actual patients. The number of participating therapists was limited. The presence of an observer might have influenced the therapists, and the effect of each individual item and the interaction between multiple changes in user interface was not systematically studied in a controlled manner. Still, using a mixed methodology approach including interviews, field observations, questionnaires, and controlled observations, an in-depth understanding was established. Reflecting on the various design iterations over these nine years provides a clear main conclusion. The trend is towards a *treatment-oriented* therapist user interface away from a *simulation-oriented* user interface. It is important to recognise the dual tasks therapists have to perform: treating the patient and creating a VR exposure. Ideally the system should be responsible for creating a realistic exposure, while the therapists only need to control the elements in the VR world that evoke anxiety.

The paper's contribution can be summarised with the following set of new design guidelines that were derived from the evaluations.

Guideline 1. Provide therapists with automated exposure scenario(s).

Guideline 2. Provide therapists with an integrated timeline representation of the different phases in the scenario, the recorded anxiety scores, comment flags, the events to come and those already taken place, and the current position on the timeline.

Guideline 3. Design for error prevention by not allowing therapists to trigger inappropriate simulation events.

Guideline 4. Provide therapists with predefined comment flags to record events in the session.

Guideline 5. In a VRET environment where the position of the patient is fixed in the VR world, therapists do not need an external viewpoint of the VR world with a projection of the patient (exception to Schuemie's (2003) general guideline that recommends a VR projection of the patient in VR world).

Although these guidelines were developed in the context of treating patients for their fear of flying, they also seem useful for VRET systems designed for treating other phobias as long as the exposure represents the unfolding of a story line. For social phobia a story line could be where patients enter a restaurant where they ask a waiter for a table, order food and compliment the waiter or complain about the food. For acrophobia a story line could simply be the path patients follow by climbing various stairs in a building.

This study also provides some more general insights when it comes to the design methodology and the area of virtual reality rehabilitation and therapy. First of all, the findings support the conclusion drawn in a SWOT (Strength, Weaknesses, Opportunities, and Threats) analysis (Rizzo & Kim, 2005) of this area. Care should be placed on building easy to use systems for therapists, going beyond system development for research purposes only. Still this might not be enough. Confidence in using the system seems essential for system acceptance in clinics. In this study this was provided by a technical assistant attending a number of therapy sessions. As Rizzo and Kim (2005) pointed out, patient-therapist relationship is seen as a key factor of a therapy. They identified the perception of VR replacing the therapist as a threat in their analysis. Although this was not directly observed in this study, the request for a technical assistant suggests that therapists need confidence in the system. If they perceive the system as unreliable or difficult to work with, they might see this technology as a threat to the patient-therapist relationship. They might fear that patients perceive them as unable to control the situation. Besides this negative perception, VR has also been suggested to contribute to the therapist-patient relationship (Riva, 2005) as it creates an environment in which patients feel more secure, enabling them to express thoughts and feelings that are otherwise difficult to discuss and thereby strengthening the therapist-patient relationship. Still, to overcome the negative perception, therapists might also need training to improve their confidence and establish trust in the system. This last observation might be typical to mental health or indeed to the health domain in general. Systems in these domains have multiple users: health providers and patients. Because of the dissimilarity in knowledge, responsibility is also placed upon the health provider to select appropriate treatment technology. In addition, health funding-organisations, such as insurers or governments, might place additional demands. In short, designing for the mental health domain seems distinctly different than for example designing for the consumer electronics domain with its own set of actors such as the consumers, family members, social peer groups, but also content providers. Furthermore, in the mental health domain opportunities to collect field observation data is more limited because of patient-therapist confidentiality, and ethical concerns about the ability of patients with mental health problems to give research consent disclosure and the way in which this should be obtained (Christopher, Foti, Roy-Bujnowski, & Appelbaum, 2007; Coyle, Doherty, Matthews, & Sharry, 2007; Stiles, Poythress, Hall, Falkenbach, & Williams, 2001).

With regard to the design methodology, this study also gives an insight into the different types of data collected. Compared to laboratory evaluations, the field observation provided a more in-depth understanding of the context the therapist is working in. For example, already in the evaluation of the first redesign the note-taking feature was not well received. However, only in the field observation it became apparent why. As with contextual inquiry (Beyer & Holtzblatt, 1998) this was obtained through a combination of field observation and establishing an interpretation together with the therapist. Furthermore, the study shows that even after a complete redesign cycle, usability improvements could still be made in a second redesign cycle. This supports, therefore, iterative design methodologies or questions the ability to solicit sound system requirements in an initial phase. One key difference between the first and the second redesign cycle was the practical experience therapists had obtained with the technology. This might put them in a better position to understand the implications of the technology. Additionally, the designer might also be a factor as additional design cycles or simply time might be needed to shape potential design solutions. In cases with new technologies, such as VR, technology is also a key driver determining the generation of design solutions. This is often combined with explicit interplay between technology and human factors, in this case for example controlling the simulation and the workload of the therapists. In these conditions the recently suggested situated cognitive engineering approach (Neerincx & Lindenberg, 2008; Neerincx, et al., 2008) might be very applicable to design for this complex tasks environments. This approach also emphasises the use of scenarios and claim analysis (Carroll, 2000) to establish the initial baseline requirements. Whereas in this study the redesign was started with a review of an already existing system, recent research explorations (Paping, Brinkman, & van der Mast, in press) have used filmed scenarios of non existing systems to discuss with therapists potential design solutions and implications. Especially in conditions where development investments are high, this technique can help to consider potential design implications before actual development, after which design modifications are often more difficult to establish. For example, by reducing the therapist's task load for controlling the VRET system, it might be possible that the therapist can be involved in other tasks. Using the situated cognitive engineering approach, recent research (Paping, et al., in press) has been exploring whether a single therapist would be able to treat multiple patients simultaneously. As patients are physically separated from each other, some form of tele-treatment seems needed. The current technical architecture with a separation of therapist and patient user interface would allow this (van der Mast, 2006). Presenting therapists with films depicting several use scenarios of a tele-VRET system has already resulted in an initial interaction design for such a system. Computer-based patient monitoring e.g. SUD score obtained with speech recognition, and physiological measures, would allow the system to support therapists to deal with multiple patients that simultaneously need attention. Based on the concept of adaptive automation (de Greef & Arciszewski, 2008), the system can temporarily take over one patient and reduce or increase the number of anxiety provoking elements in the patient's VR world or switch to a completely other scenario. Another future role the computer can play is to simulate patients as part of a VRET training environment for student therapists. Students could be trained in running a session and by using virtual characters that can express emotions, students can also learn to monitor the anxiety of a patient. Still in all these new challenges a key focus will remain the usability of the therapist user interface. Designers should keep in mind that therapist's main goal is to cure the patient and not to control the VR simulation.

Appendix A

Therapist user interface usability questionnaire

Please indicate with a check mark from -3 (completely disagree) to + 3 (completely agree) whether you agree with the following statements.

No	Statement
1	I had quickly learned how to use the system
2	I found the system to be easy to use.
3	I (subjectively) like to use the system
4	I found it easy to control the virtual world using the interface given.
5	It was easy to control what I wanted to do
6	I could do all things that I wanted to do
7	It was immediately clear what I could and what I couldn't do in the system
8	I felt I was in complete control during the session
9	I like using the controls of the system
10**	I found the error message is easy to understand

No	Statement
11	I found it was easy to correct the mistake I have made
12*	Sometimes I feel I lost my orientation in controlling the Virtual World, like often forget 'In which stage I am' or 'When this stage will be over'
13	I could estimate how long the session will last and could plan the session precisely
14	Language used is easy to understand and to memorized
15	I found it easy to fill the session and patient Information
16	I found the map to be clear and unambiguous
17	I found the flight control to be easy to use
18**	I could see what I wanted to see on the patient's viewpoint
19**	I could see what I wanted to see on the free viewpoint
20**	I found the controls of the free viewpoint easy to use
21**	I found it easy to use the cloud control
22	I found the SUDs recording is easy to use
23**	I found the size of SUD chart size big enough
24	I found the alarm clock easy to use
25	It was easy to control the voice of pilot and purser
26**	I found the Notes size reasonable
27	I found the user interface is used efficiently.

**the score for this question was reversed by multiplying it with –1; **not included in evaluation of second redesign*

Appendix B

Therapist's user interface questionnaire for features added in first redesign

Please rate these statements on a scale from -3 (completely disagree) to +3 (completely agree)

No	Statement
1	I found it easy to control the Flight plan
2	I found the Cabin Control is easy to use
3	I found the Roll Control is easy to use
4	I found the Flight View is very helpful
5	I found the Print Function very useful
6	I found the report easy to understand
7	The timer provided much help
8	I found it easy to start and end the session

Appendix C

Real therapist only questionnaire

Please rate these statements on a scale from 1 (completely disagree) to 5 (completely agree)

No	Component	Statement
1	Roll controls	I often used the roll controls
2		I found the roll controls to be easy to use
3	Bad weather control	I often used the bad weather control
4		I found the bad weather control to be easy to use
5	Flight view	I often use the flight view
6		The flight view give me an overview of the session
7		The timer given is very helpful
8		In the therapist control condition, I have overview over the whole situation

No	Component	Statement
9	Report	The report feature is very useful

References

APA (2000). *Diagnostic and statistical manual of mental disorders* (4th, text revision ed.). Washington, DC: American Psychiatric Association.

Banos, R. M., Botella, C., Perpina, C., Alcaniz, M., Lozano, J. A., Osma, J., et al. (2002). Virtual reality treatment of flying phobia. *IEEE Trans Inf Technol Biomed, 6*(3), 206-212.

Bentley, T. (2000). *Biasing web site user evaluation: A study*. Paper presented at the OZCHI.

Beyer, H., & Holtzblatt, K. (1998). *Contextual design: Defining customer-centered systems*. San Francisco, CA: Morgan Keufmann.

Brinkman, W. P., Haakma, R., & Bouwhuis, D. G. (2009). The theoretical foundation and validity of a component-based usability questionnaire. *Behaviour & Information Technology, 28*(2), 121-137.

Brinkman, W. P., Sandino, G., & van der Mast, C. (2009). *Field observations of therapists conducting virtual reality exposure treatment for the fear of flying*. Paper presented at the ECCE 2009.

Brinkman, W. P., van der Mast, C. A. P. G., & de Vliegher, D. (2008). *Virtual reality exposure therapy for social phobia: A pilot study in evoking fear in a virtual world*. Paper presented at the HCI2008 workshop - HCI for technology enhanced learning.

Carroll, J. M. (2000). *Making Use: Scenario-based Design of Human-Computer Interactions*. Cambridge: MIT Press.

Christopher, P. P., Foti, M. E., Roy-Bujnowski, K., & Appelbaum, P. S. (2007). Consent form readability and educational levels of potential participants in mental health research. *Psychiatric Services, 58*(2), 227-232.

Coyle, D., Doherty, G., Matthews, M., & Sharry, J. (2007). Computer in talk-based mental health interventions. *Interacting with Computers, 19*, 545 - 562.

Curtis, G. C., Magee, W. J., Eaton, W. W., Wittchen, H. U., & Kessler, R. C. (1998). Specific fears and phobias. Epidemiology and classification. *Br J Psychiatry, 173*, 212-217.

da Costa, R. T., Sardinha, A., & Nardi, A. E. (2008). Virtual reality exposure in the treatment of fear of flying. *Aviat Space Environ Med, 79*(9), 899-903.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *Mis Quarterly, 13*(3), 319-340.

de Greef, T., & Arciszewski, H. (2008). *Combining adaptive automation and adaptive teams in a naval command centre*. Paper presented at the European conference on cognitive ergonomics.

Emmelkamp, P. M. G. (2004). Behaviour therapy in adults. In M. Lambert (Ed.), *Bergin and Garfield's Handbook of psychotherapy and behavior change* (5 ed., pp. 393 - 446). New York, NY: Wiley.

Garcia-Palacios, A., Botella, C., Hoffman, H., & Fabregat, S. (2007). Comparing acceptance and refusal rates of virtual reality exposure vs. in vivo exposure by patients with specific phobias. *Cyberpsychol Behav, 10*(5), 722-724.

Gregg, L., & Tarrier, N. (2007). Virtual reality in mental health : a review of the literature. *Soc Psychiatry Psychiatr Epidemiol, 42*(5), 343-354.

Grillon, H., Riquier, F., Herbelin, B., & Thalmann, D. (2006). Virtual reality as a therapeutic tool in the confines of social anxiety disoder treatment. *International journal on disability and human development, 5*(3), 243-250.

Gunawan, L. T., van der Mast, C., Neerincx, M. A., Emmelkamp, P., & Krijn, M. (2004). *Usability of therapist's user interface in virtual reality exposure therapy for fear of flying*. Paper presented at the 10th Annual Scientific Conference on Web Technology, New Media, Communications and Telematics Theory, Methods, Tools and Applications.

Krijn, M., Emmelkamp, P. M., Biemond, R., de Wilde de Ligny, C., Schuemie, M. J., & van der Mast, C. A. (2004). Treatment of acrophobia in virtual reality: the role of immersion and presence. *Behav Res Ther, 42*(2), 229-239.

Krijn, M., Emmelkamp, P. M., Olafsson, R. P., & Biemond, R. (2004). Virtual reality exposure therapy of anxiety disorders: a review. *Clin Psychol Rev, 24*(3), 259-281.

Krijn, M., Emmelkamp, P. M., Olafsson, R. P., Bouwman, M., van Gerwen, L. J., Spinhoven, P., et al. (2007). Fear of flying treatment methods: virtual reality exposure vs. cognitive behavioral therapy. *Aviat Space Environ Med, 78*(2), 121-128.

Loewenthal, K. M. (2001). *An introduction to psychological tests and scales*. New York, NY: Psychology press.

Muhlberger, A., Herrmann, M. J., Wiedemann, G. C., Ellgring, H., & Pauli, P. (2001). Repeated exposure of flight phobics to flights in virtual reality. *Behav Res Ther, 39*(9), 1033-1050.

Neerincx, M. A., & Lindenberg, J. (2008). Situated cognitive engineering for complext task environments. In J. M. C. Schraagen, L. Militello, T. Ormerod & R. Lipshitz (Eds.), *Naturalistic decision making and macrocognition* (pp. 373 - 390). Aldershot, UK: Ashagte.

Neerincx, M. A., te Brake, G. M., van de Ven, J. G. M., Arciszewski, H. F. R., de Greef, T. E., & Lindenberg, J. (2008). *Situated cognitive engineering: Developing adaptive track handling support for naval command and control centers*. Paper presented at the HCP-2008 - Third International Conference on Human-Centred Processes.

Nielsen, J. (Ed.). (2002). *Coordinating user interfaces for consistency*. San Francisco, CA: Morgan Kaufmann.

Nielsen, J., & Molich, R. (1990). *Heuristic evaluation of user interfaces*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people.

Norman, D. A. (1981). Categorization of action slips. *Psychological review, 88*(1), 1-15.

Paping, C., Brinkman, W. P., & van der Mast, C. (in press). An explorative study into tele-delivered multi-patient virtual reality exposure therapy system *Wounds of War II: Addressing posttraumatic stress disorder in peacekeeping and combat troops*: IOS press.

Parsons, T. D., & Rizzo, A. A. (2008). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis. *J Behav Ther Exp Psychiatry, 39*(3), 250-261.

Powers, M. B., & Emmelkamp, P. M. (2008). Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *J Anxiety Disord, 22*(3), 561-569.

Riva, G. (2005). Virtual reality in psychotherapy: Review. *Cyberpsychology & Behavior, 8*(3), 220-230.

Rizzo, A., & Kim, G. J. (2005). A SWOT analysis of the field of virtual reality rehabilitation and therapy. *Presence-Teleoperators and Virtual Environments, 14*(2), 119-146.

Robson, C. (2002). *Real world research* (2 ed.). Malden, MA: Blackwell.

Schuemie, M. J. (2003). *Human-computer interaction and presence in virtual reality exposure therapy.* Delft University of Technology, Delft, The Netherlands.

Schuemie, M. J., Abel, B., van der Mast, C. A. P. G., Krijn, M., & Emmelkamp, P. M. G. (2005). The effect of locomotion technique on presence, fear and usability in a virtual environment. *EUROMEDIA '2005: 11th Annual Euromedia Conference*, 129-135.

Stiles, P. G., Poythress, N. G., Hall, A., Falkenbach, D., & Williams, R. (2001). Improving understanding of research consent disclosures among persons with mental illness. *Psychiatric Services, 52*(6), 780-785.

van der Mast, C. A. (2006). Technological challenges in the Delft virtual reality exposure therapy system. *International journal on disability and human development, 5*(3), 205-210.

Wiersma, J., Greeven, A., Berretty, E., Krijn, E., & Emmelkamp, P. (2008). De effectiviteit van Virtual Reality Exposure Therapy voor hoogtevrees in de klinische praktijk. *Gedragstherapie, 41*, 253-259.

Wolpe, J. (1958). *Psychotherapy by reciprocal inhibition*. Stanford: Stanford University Press.