

Developing a Dialogue Editor to Script Interaction between Virtual Characters and Social Phobic Patient

Niels ter Heijden, Chao Qu, Pascal Wiggers, Willem-Paul Brinkman
Delft University of Technology
The Netherlands

slein13@gmail.com, aquchaos@gmail.com, p.wiggers@tudelft.nl, w.p.brinkman@tudelft.nl

ABSTRACT

Virtual Reality Exposure Therapy (VRET) has been put forward as a treatment for patient suffering from anxiety disorder such as social phobia. Current VRET systems however provide limited speech interaction possibilities between the patient and virtual characters and therefore does not seem to offer patients an exposure to the full richness of an actual human-human dialogue. One way to support a free speech dialogue between a patient and a virtual character is to develop interactive pre-scripted dialogue scripts, where specific patient answers can trigger pre-recorded avatar responses thereby creating extensive dialogue trees. This paper discusses this approach and a dialogue editor to write these dialogue scripts. Online chat bots are proposed as a technique to evaluate and to improve an interactive dialogue script. Results of a pilot study with 4 non-phobic individuals are promising and suggest that these scripted interactive dialogues can be used to simulate a human-human dialogue.

Categories and Subject Descriptors

H.5.1 [Multimedia Information Systems] virtual realities

General Terms

Design, Experimentation, Human Factors.

Keywords

social phobia, keywords, dialogue, conversation, database, avatar.

INTRODUCTION

Social phobia is one of the most common types of anxiety disorders. People with a social phobia have a strong fear of social situations, such as talking in public, entering a room with other people, ordering food in a restaurant etc. Social phobia is associated with depression, substance abuse (e.g. alcoholism, drug abuse), restricted socialisation, and poor employment and education performance [8,9]. In the western world, social phobia leads to intensive use of (mental) health services [20]. When persons with social fears seek professional help, they do it most often after a long period of increasing complaints (on average 15 year) [9] and are being treated with exposure in vivo (i.e. exposure to actual real-life situations). Although effective, this treatment has a number of serious drawbacks such as high costs, drop outs, and fundamental constrains in the scope, control and duration of the exposure. Exposure in virtual reality has been suggested as alternative for these

limitations. Virtual reality exposure therapy has been shown to be effective for other phobias such as fear of flying and fear of heights. However, current virtual reality exposure therapy systems do not seem to offer an extensive simulation of human-human dialogues, which would greatly enhance the applicability of this technique in the treatment of social phobia. The suggestion offered in this paper is to use interactive dialogue scripts to, allow a patient to talk with a virtual character in a Virtual Environment (VE). As the design of virtual worlds need their own set of design tools, the design of these interactive dialogue scripts also need support of tools such as a dialogue editor, especially as these scripts rapidly become vast and complex to manage. Although naturally behaving virtual characters, i.e. avatars, seems an important element in establishing high level of presence, a fully free natural speech processing system is not yet feasible [7]. For this reason, a robust dialogue system is developed that aims at creating a natural and stable automatic human-like conversation in virtual environment. It is based on question - answer structure of a conversation, making use of keyword detection to select pre-defined response from a database.

Social Phobia and Treatment

Social anxiety is defined as anxiety about social situations, interactions with others, and being evaluated or scrutinized by other people [13]. It includes emotional discomfort, fear, apprehension, or worry. As one of the social anxiety category, social phobia is a common occurring anxiety disorders, estimated to affect 13.3% of the US population [10]. Patients suffered by social phobia will have a serious impact on the life, education, career and social life. Several different kinds of treatments are in use to deal with social phobia. Cognitive Behaviour Treatment (CBT) is one of those treatments for which the effectiveness has been demonstrated [6]. An important element of CBT in treatment of a phobia is to put the patient in the feared situation after therapist discussed the fear level and provides the patient with another perspective or relaxation techniques. Patients are gradually exposed to the feared situation; they progress to more anxiety evoking situations once they get habituated to the situation and anxiety subsides. Choosing a correct situation for patients and convincing them to face it is the most difficult part of this kind of CBT treatment.

VRET

Using a VE makes it possible to expose a patient to the feared situation in a computer generated environment. It shifts the place of exposure from real world to virtual reality. With the ability to control all objects in the VE, VRET is a tool to ease the difficulty in putting the patient in the feared situation. It can easily generate a room with a varying number of avatars, or for example place patients in a train station, a restaurant or in a shop. VRET systems for several phobias that can be treated with CBT have been developed. These phobias include claustrophobia, fear of driving, acrophobia, spider phobia, panic disorder with agoraphobia, post traumatic stress disorder, fear of flying and also social phobia, especially simulations of public speaking situations [1]. Recreating also other social senses have been tried out by Klinger [11,12]. Meta-studies [5,16,17] on the effectiveness VRET are encouraging, suggesting that the treatment is as effective as exposure in vivo.

The research and development of VRET system for social phobia is still in process. Studies until now focused on fairly static virtual characters (avatars) some made with real photographs [12], others using low detailed computer generated images [19]. Mostly the interaction between the patient and avatars is limited to a specific body language the avatars conveyed towards the user (e.g. attentive audience, or audience that is bored). As far as we know, there are no reports in the literature of fully automated speech interaction between the avatars and patient simulating a real conversation a patient would otherwise have with for example a shop assistant. Having this ability however would extend the range of possible types of exposures that could be offered, and automating part the simulation would reduce the workload of the therapist, as has also been suggested in VRET systems to treat patient with a fear of flying [3].

VRET for Social Phobia

To treat social phobia through VRET, the patient – avatar communication is a key component, especially if it is possible to have the patient experience a sufficient level of presence thereby perceive the interaction as lifelike. However, letting a computer recognize completely free speech is too ambitious to realize with existing technology [7,14]. Instead semi-automatic alternatives seem more feasible. For example, therapists could control most of the avatars' behaviours, by simply listing to patient and select appropriate responses. In this way, avatars might have a realistic (verbal) behaviour; the drawback however is the extensive workload forced up on the therapists. An alternative approach, which removes much of this workload, is to have patients read out loud one of the patient-responses from a short list of responses shown on the screen [2]. This method can fairly successful be implemented with speech recognition. But with a list of sentence on the screen it is unclear how this might affect the level of presence and the level of anxiety VE might still be able to evoke. Instead, an intermediating solution is put forward which uses automatic keyword detection from the patients free speech.[15] Using a dialogue system and with the help of automatic speech recognition software, the system would

get verbal actions of the patient and select automatically from database of pre-recorded avatar responses. For different scenarios and topics, the dialogue system only needs to be changed to the relevant database. This semi-automatic solution would therefore reduce the workload of the therapists and establish sufficient level of presence.

Question Creation

The success of this approach depends on the careful selection of questions and appropriate avatar responses on the patient's verbal actions. To enhance the dialogue creation process the following categories are suggested to classify possible patient actions:

- A. *keyword* class; there are pre-defined keywords in patients' answers.
- B. *yes/no* class; the answer is simply 'yes' or 'no'.
- C. *attitude* class; the patients put up positive or negative attitude in their replies.
- D. *unknown* class; the patients do not know or uncertain about the questions.
- E. *length limited* class; the answer is consisted by limited words. The limitation length is a pre-defined number.
- F. *general* class: Any answers which could not be classified in the above classes are treated in this class.

Each category can be linked with its own predefined computer response. In other words, for each question the avatar brings forward to the patient, the correspondent avatar-response on the patient's answer should be stored in the dialogue database in advance. This allows the system to choose a suitable response from the answer category. Priority levels should be assigned to each class if patient reaction applied to multiple answer categories. When writing an interactive dialogue script, the following four points might be considered:

1. When selecting a conversation topic, more neutral, time independent, or topics with well established social scripts are preferable (e.g. ice age, gardening, recycling) If the topic is too sensitive (e.g. sexuality, religion, or childhood), the patients might give answers fully immersed with their emotional colouring, specific point of view and their experience. As these answers might vary greatly, it will be more difficult to anticipate appropriate avatar replies.
2. Include open questions to invite patient to give more elaborated response instead of simply 'yes', 'no', or 'don't know'. This might help creating longer dialogue and extent duration of the exposure.
3. Pose questions that will result in answers with detectable keywords. Being able to pose related sub questions might give patient a feeling that the avatar is really listening to their answers.
4. Although there are several answer categories, it is not necessary to include all categories for each

question. In some situations, only including a general response might be satisfactory.

DIALOGUE TOOL

The first step to develop an interactive dialogue script is to create a dialogue tree. Already a conversations tree with a depth of 4 or more answer-responses branches might become too complex to manage without the support of an software tool. The dialogue tool Editor3 has especially been created to support dialogue designers with creating interactive dialogue scripts. The tool gives an overview of what has already been written and the structure of the dialogue. It also allows the reuse of dialogue parts and the ability to merge dialogue branches back together again. The dialogue tree needs to be stored and appended with extra meta-information. This meta-information is read and used by the eventual program which reads the stored interactive dialogue script and controls the avatars.

Database structure

Editor3 provides a visual way to fill in a SQLite database containing the interactive dialogue script. The most basic elements of a dialogue tree are the sentences and the links between them. The database therefore represents these two elements. The order of the links of one sentence to another can also hold some importance or meaning, for example the first sentence would be a positive reaction while the second a negative one. So the order of the links is also represented in the database. Every sentence can hold some meta-data such as which avatar has to say the sentence if more than one avatar is involved or more basic, the name of the audio file that should be played or an animation that should be played along with the audio file. Because there is a visual component to the final reproduction of the dialogue in the shape of a virtual world with actors, it might be necessary to do some actions before or after a sentence has been said. Such as an avatar standing up and walking to another place in the virtual world. This also can be represented in the database so that a complete scenario can be created in one database.

Editor3

As mentioned before, Editor3 is created to fill the database in an easy visual way. The tool provides all the functionality the database can represent and provide the dialogue designer with some extra functionality. The program starts up with an empty root sentence which is the beginning of the entire dialogue tree. This root sentence can then be linked to new empty sentences which create the first branched of the tree (Figure 1). Linking can also be done to existing sentences in the tree hereby providing a way to merge branches back together. Every sentence can be annotated with extra meta-data in the form of properties. A property has a name which is equal over all sentences and a value which can vary for each sentence. Editor3 provides a way to create new properties or selects an existing one. The property value is a string. The specific meaning to a property must be programmed in the application that later on controls the avatars and interpreter the database. The same applies for the links and the selection of the appropriate follow up

sentence for the dialogue. Finally every sentence can have a list of ordered pre and post actions that need to be executed before or after the sentence is said.

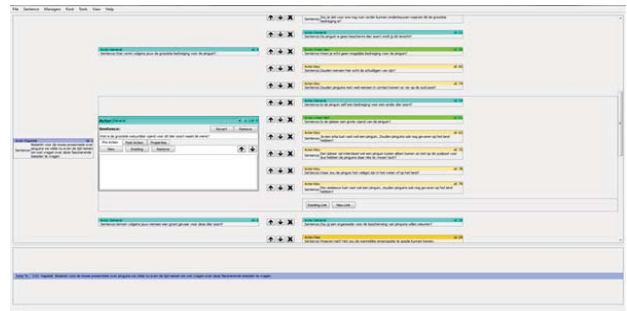


Figure 1: Editor3 with editable sentence in the centre

When adding sentences and links to the database, designers can of course make mistakes or change their mind. The editor therefore support the deletion of links and sentences form the database. This has some implications because removing a sentence will leave links toward that sentence empty and can create orphan tree parts. Editor3 therefore provides a more advance remove function that will clean up after the removed sentence or link.

Each sentence gets by default the Actor property. This property contains the name of the avatar that needs to speak the sentence (or in other cases what differentiate the sentence from the other linked sentences). Because of the special meaning of this property it is always shown in the editor even if the other properties are not shown.

Editor3 also provides the dialogue designer with some additional functions to support the design process such as a print function that can print out all possible conversation routes through the tree or let dialogue designers select their own route through the tree. Editor3 also provides a tool to record sound from a microphone and save it to a directory while adding the filename to a property of a sentence. In this way designers can systematically record all sentences and linking them to the database.

When editing sentences or links between them, the tool only shows a limited part of the dialogue tree, consisting out of the previous sentence, the sentence that is being edited and all the other linked sentences to the previous sentence and all the sentences linked to these sentences (Figure 1). This produces three columns one for the previous sentence, one for all current sentences that can be edited and all the next sentences. Clicking on the previous or next sentences will jump the current column to the clicked sentence column and make the clicked sentence editable. It allows the dialogue designer to navigate through the dialogue tree. This way of showing the dialogue tree was chosen to limit the information on the screen to the absolute necessary information. But even now, with some wide branching dialogue trees, the information can not be shown onto one screen and scrolling is still needed. Finally the tool provides a non-editable view of the entire dialogue tree (Figure 2) that with even simple dialogues already needs a lot of

scrolling, but might give dialogue designers some sense of the overall structure of dialogue.

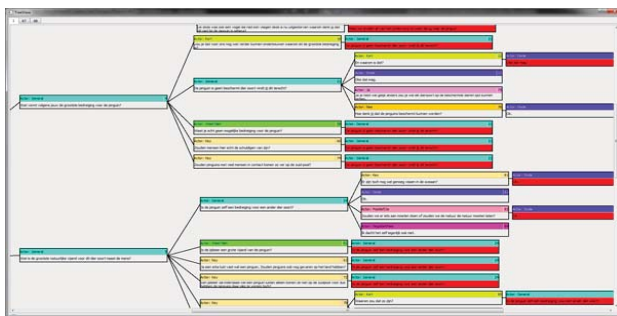


Figure 2: Editor3 TreeView of an actual dialogue. All red sentences are merges to the main story branch

Example Dialogue

Figure 3 shows a tree view of a dialogue tree about penguins in Dutch, and a written translation into English is given below.

[Actor: General]
 What do you think is the main threat to penguins?

- [Actor: Short]
 Could you further elaborate on why that would be the main threat?
- [Actor: General]
 The penguin is not a protected animal, do you think that is right?
- [Actor: Don't know]
 Do you really not know of any threat to penguins?
- [Actor: Key]
 Would humans be the real perpetrator of this?
- [Actor: Key]
 Would penguin come much in contact with humans so far away on the South Pole?

The Actor property shows when the linked sentence should be chosen by the interpreter. In this example possible reasons can be that: patients give only a short response, patients say yes or no, patients respond negatively or positively, patients respond that they do not know the answer, or patients say specific keywords. Figure 3 is a close up of Figure 2 where a General response sentence is linked to five possible follow up responses of avatar on the patient's interaction that occurs in between the sentences shown. This dialogue tree is a final product of multiple iterations and was used in a pilot study discussed later on.



Figure 3: Close up of TreeView

CHAT BOT EVALUATION

To explore the feasibility of the dialogue editor, we created four interactive dialogue scripts. All scripts were written for a social scenario in which a patient had to talk for a number of minutes about a topic after which four avatars would ask the patient a series of questions on this topic. As the target group was very generic, Dutch adults, four fairly neutral topics were selected: democracy, France, dogs and penguins. To limit the complexity of the dialogue, the interactive dialogue script did not anticipate questions from the patients. Development of robust dialogues is best done in an iterative evaluation cycle [7]. To evaluate and to identify more keywords for the four dialogues a simple online MSN chatbot was created. The chatbot had as advantage that it did not use a real speech recognizer and therefore did not require training the recognizer and avoided the unavoidable recognition errors. Also the participants could participate from their home on their own time. Still, a possible negative effect using the chatbot could be that people respond differently when writing their response instead of saying them. The only extra factor to take into account with the chatbot was spelling errors on the side of the participant. These were addressed by adding alternative spellings of the keywords. Two chatbot iterations were conducted, and started without any initial keywords. The observation of the first iteration was the need for “don't know” answer. The answers categories short, yes/no and attitude where found adequate and only needed some slight adjustments. The changes to the dialogue tree as a whole was still extensive enough to justify a second chatbot iteration especially the addition of “don't know” sentences needed testing. Table 1 shows the data of the second iteration¹⁹.

PILOT STUDY

The pilot study consisted out of four individuals. None of them suffered from social phobia. They completed all four dialogues with four different speech recognition conditions (human controller, speech length detection, simple speech recognition, and full keyword detection). For this paper only human controller and full keyword detection will be discussed.

Setup

The participants were guided into the room where the experiment would take place and seated before a table with a microphone and some papers. First the experiment was explained to them and what data was collected and used. Also the use of a camera for remote observation was explained. After this short introduction the participants were asked to train the commercial speech recognition tool (Nuance Naturally Speaking). The short training took around 5 to 10 minutes. After the training of the speech recognizer the experiment was explained in more detail. In each of the four dialogues the participants were asked to give a short presentation about the topic.

¹⁹ Note that the France dialogue was only evaluated in a single iteration cycle and started with an initial keywords set

Table 5: Start data and results of chat bot evaluation

Dialogue	Number of starting questions	Mean depth of three	Initial number of keywords	Number of keywords afterwards	Number of participants	Total Number of keywords hits	Mean (SD) number of user turns
Democracy	9	3.1	2	4	10	2	18.0 (2.9)
France	8	2.5	11	18	10	32	18.6 (1.1)
Dogs	8	1.9	8	12	5	8	13.2 (2.1)
Penguins	10	2.6	11	16	6	5	21.0 (2.0)

To help them with the presentation, a paper with some information bullet points about the topic was provided. Before every presentation the participant was given time to study this paper. The points on the paper were chosen to not address questions asked by the avatars later on. The participants were told that after each presentation the avatars would ask some questions about the topic. The participant was asked to answer these questions normally and as good as they could. After each question round the participant was asked to fill in a questionnaire how they had experienced the question round. After the explanation participants were asked to fill out a background information questionnaire and the Personal Report of Confidence as a Public Speaker (PRCS) [4] questionnaire. While the forms were filled out, the virtual world was started up and appeared on a projection screen in front of them (Figure 4). The projection was wall size leaving only half a meter to the floor without projection. After filling out the two forms the participant was told that all sound would be heard over the headphones.



Figure 4: Virtual room with four avatars that was projected on an wall

Next, the participant was left alone in the room and given some time to read the bullet point paper. The experiment was started as soon as the participant signalled that he was ready. After doing all four

dialogues with presentation, question and filling in the questionnaire a final questionnaire was handed over. The participants were told that this IPQ questionnaire [18] should capture the entire scope of the four experiments.

Result

Although the sample size was too small for quantitative analysis, some general remarks can already be made about effectiveness of the set up. As expected, the participants tend to give longer and more detailed answers when using speech to communicate. Nonetheless the keywords used in the answers seem to match those used with the chat bot. This suggests that the reactions from the computer characters on things said by the user seems appropriate. When the participants were asked what gave them the feeling of presence the precise responses from the avatars on what they had said was often mentioned. Because people tend to talk longer the responses on short answers were rarely used. All the other responses given by the computer seemed appropriate to what the participants said. Only in rare cases especially in responses from the participant with negative and positive words the computer response was completely wrong. In all other cases no reaction from the participants was noted that suggested that they felt the response was weird or wrong.

CONCLUSION AND DISCUSSION

Although limited in size, the observations of pilot study seems to suggest that it might be possible to create a VE simulation of human-human dialogue. Furthermore, Editor3 and the online chatbot evaluation seem promising tools in creating interactive dialogue script. Still, future work is needed to validate these preliminary observations. Especially, studying with people that suffer from social phobia seems necessary. They might give shorter answers or different answers. Also adding a response for long answers would be interesting because people now felt that the computer responded uninterested on their elaborate answers. Also the handling of questions from the individuals would be interesting. Now the participants were asked not to ask the avatars questions. How well would patients follow this instruction? Finally the computer should be able to

handle request to repeat its last sentence. In case an avatar's question was not fully understood or heard by the participant. Exploring these issues will help in creating a VRET system that allows social phobic patient to be exposed to social scenarios that include an important social element, which is talking with people.

REFERENCES

- [1] Anderson, P. L., Zimand, E., Hodges, L. E., & Rothbaum, B. O. (2005). Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and Anxiety*, 22(3), 156-158.
- [2] Brinkman, W.-P., Mast, v. d., C.A.P.G., & de Vliegheer, D. (2008). Virtual reality exposure therapy for social phobia: A pilot study in evoking fear in a virtual world. *Proceedings of HCI2008 workshop - HCI for technology enhanced learning*, 85-88.
- [3] Brinkman, W.-P., van der Mast, C., Sandino, G., L.T., G., & Emmelkamp, P. (2010). The therapist user interface of a virtual reality exposure therapy system in the treatment of fear of flying. *Interacting with Computers*, 22(4), 299-310.
- [4] Gilkinson, H. (1942). Social fears reported by students in college speech classes. *Speech Monographs*, 9, 131-160.
- [5] Gregg, L., & TARRIER, N. (2007). Virtual reality in mental health - A review of the literature. *Social Psychiatry and Psychiatric Epidemiology*, 42(5), 343-354.
- [6] Hofmann, S. G., & Barlow, D. H. (1999). The costs of anxiety disorders: Implications for psychosocial interventions. *Cost-effectiveness of psychotherapy*, 224-234.
- [7] Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition* (2nd ed.). Upper Saddle River, N.J.: Pearson Prentice Hall.
- [8] KATZELNICK, D. J., KOBAK, K. A., DeLeire, T., Henk, H. J., Greist, J. H., Davidson, J. R. T., et al. (2001). Impact of generalized social anxiety disorder in managed care. *American Journal of Psychiatry*, 158(12), 1999-2007.
- [9] Kessler, R. C. (2003). The impairments caused by social phobia in the general population: implications for intervention. *Acta Psychiatrica Scandinavica*, 108, 19-27.
- [10] Kessler, R. C., McGonagle, K. A., Zhao, S. Y., Nelson, C. B., Hughes, M., Eshleman, S., et al. (1994). Lifetime and 12-Month Prevalence of Dsm-Iii-R Psychiatric-Disorders in the United-States - Results from the National-Comorbidity-Survey. *Archives of General Psychiatry*, 51(1), 8-19.
- [11] Klinger, E., Bouchard, S., Legeron, P., Roy, S., Lauer, F., Chemin, I., et al. (2005). Virtual reality therapy versus cognitive behavior therapy for social phobia: A preliminary controlled study. *Cyberpsychology & Behavior*, 8(1), 76-88.
- [12] Klinger, E., Legeron, P., Roy, S., Chemin, I., Lauer, F., & Nugues, P. (2004). Virtual reality exposure in the treatment of social phobia. *Stud Health Technol Inform*, 99, 91-119.
- [13] Leitenberg, H. (1990). *Handbook of social and evaluation anxiety*. New York: Plenum Press.
- [14] Martin, H. V., Botella, C., Garcia-Palacios, A., & Osma, J. (2007). Virtual reality exposure in the treatment of panic disorder with agoraphobia: A case study. *Cognitive and Behavioral Practice*, 14(1), 58-69.
- [15] McTear, M. F. (2002). Spoken Dialogue Technology: Enabling the Conversational User Interface. *ACM Computing Surveys*, 34(1), 90-169.
- [16] Parsons, T. D., & Rizzo, A. A. (2008). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, 39(3), 250-261.
- [17] Powers, M. B., & Emmelkamp, P. M. (2008). Virtual reality exposure therapy for anxiety disorder: a meta-analysis. *J. Anxiety Disord*, 22(3), 561-569.
- [18] Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence-Teleoperators and Virtual Environments*, 10(3), 266-281.
- [19] Slater, M., Pertaub, D. P., & Steed, A. (1999). Public speaking in virtual reality: Facing an audience of avatars. *Ieee Computer Graphics and Applications*, 19(2), 6-9.
- [20] Wiederhold, B. K., & Wiederhold, M. D. (2005). *Virtual reality therapy for anxiety disorders : advances in evaluation and treatment* (1st ed.). Washington, DC: American Psychological Association.

Human-Computer dialogue
 Developing a Dialogue Editor to Script Interaction
 between Virtual Characters and Social Phobic Patients
 Niels ter Heijden, Chao Qu, Pascal Wiggers, Willem-Paul Brinkman
 31-10-2010

TU Delft
 Challenge the Future

Overview

- VRET
 - Social phobia
- Dialogue
 - Creation
 - Editor
 - Evaluation
- Pilot study
 - Setup
 - Results
- Experiment
 - Preliminary results

TU Delft Human-Computer Dialogue 2

1.

Virtual Reality Exposure Therapy

VRET

TU Delft Human-Computer Dialogue 3

VRET Phobias

- CBT, In Vivo
- VRET as tool for CBT
 - More control
Parsons et al., 2008
- VRET usages
 - Fear of Flying, Heights, Driving, Spiders
 - Acrophobia
 - Claustrophobia
 - PTSD
 - Social Phobias

TU Delft Human-Computer Dialogue 4

Fear of Flying **Social phobia** **Acrophobia**

all pictures from <http://www.vrphobia.com>

TU Delft Human-Computer Dialogue 5

VRET Social Phobia

- Social Phobia
 - 13% US population
 - Impact on life/career
Kessler et al., 2004
- CBT
 - Demonstrated effectiveness
Holman et al., 1999
- Existing VRET worlds
 - Photo graphic Avatars
 - Computer Animated
- Limited speech interaction

TU Delft Human-Computer Dialogue 6



E. KLINGER et al., 2005



Willem-Paul Brinkman et al., 2008

VRET Social Phobia

Focusing on Speech interaction

- Free speech
 - Therapist controls
- Scripted dialogue
 - Text on screen
- Smart scripted responses
 - Free (Directed) speech (patient)
 - Keyword detection (computer)

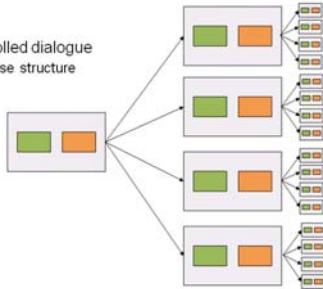


2.

Dialogue

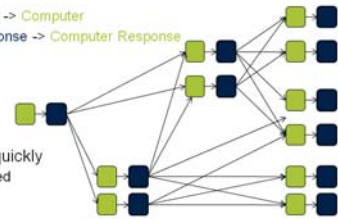
Dialogue Creation

- Structured and controlled dialogue
 - Question -> Response structure
 - Neutral topic
 - Guiding Questions
- Response groups
 - Keyword
 - Yes/No
 - Attitude
 - Unknown
 - Length
 - General
 - Question



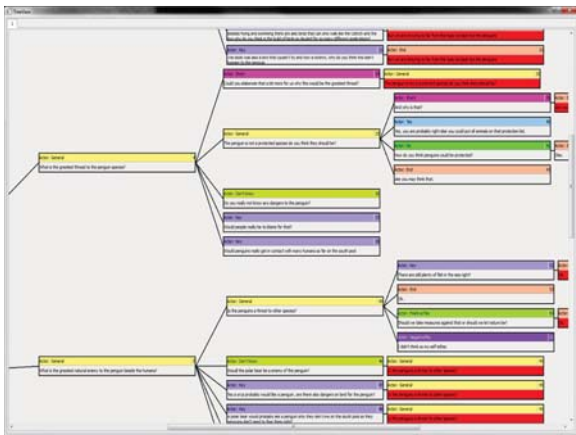
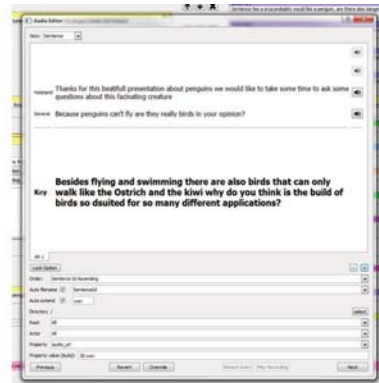
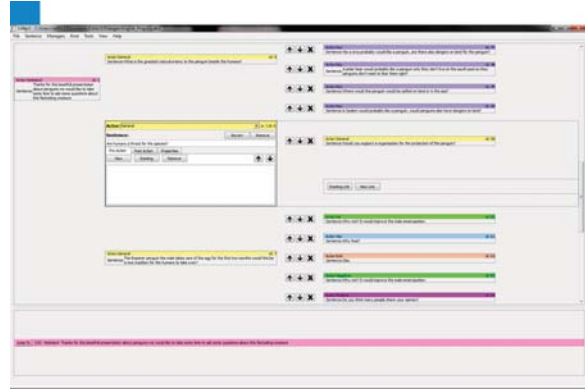
Dialogue Editor

- Dialogue tree
 - Computer -> User -> Computer
 - Question -> Response -> Computer Response
- Complexity grows quickly
 - Special tool needed



Dialogue Editor3

- Dialogue creation tool
- Editor3
 - Easy-to-use visual interface
 - Support for branch merging
 - Meta information
 - Control information
- Sound recording
- Creates SQLite database
- Only need an interpreter



Dialogue Evaluation

- Iterative design dialogue
- MSN Chatbot
- 4 topics selected
- Keywords added



Dialogue	Number of starting questions	Mean depth of three	Initial number of keywords	Number of keywords afterwards	Number of participants	Total Number of keywords hits	Mean (SD) number of user turns
Democracy	9	3.1	2	4	10	2	18.0 (2.9)
France	8	2.5	11	18	10	32	18.6 (1.1)
Dogs	8	1.9	8	12	5	8	13.2 (2.1)
Penguins	10	2.6	11	16	6	5	21.0 (2.0)

3.

Pilot Study

Pilot Study

Setup (Movie)

- 4 Participants
- 3D world projected on screen
- Participants alone in the room
- Audio interaction
- 4 Testing conditions
 - Speech detection
 - Speech recognition (limited keywords)
 - Speech recognition (keywords)
 - Human controlled



Pilot Study

Results

- Used same keywords
- Spoke longer and more elaborate
- Responses on content increased realism
- Waiting time before response decrease realism

4.

Experiment

Experiment

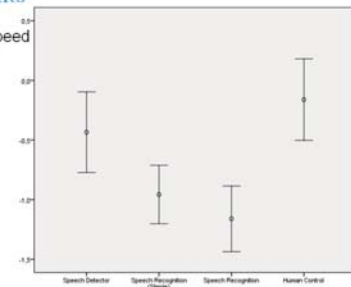
Setup / remarks

- 24 participants
 - Dutch, Adults
- Participant remarks
 - Short avatar introduction of them selves
 - To intensive staring
 - Responses on content increase presence (reported)
 - Delay in response decreased presence (reported)

Experiment

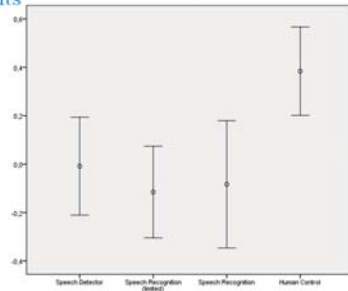
Preliminary results

- Dialog interaction speed
- $F(3,21) = 11.49$
 $p = 0.001$



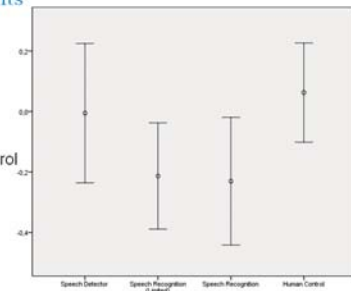
Experiment Preliminary results

- Interaction Feeling
- $F(3,21) = 11.878$
 $p = 0.001$



Experiment Preliminary results

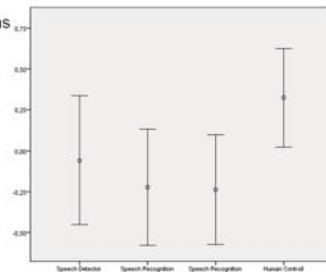
- Interaction Reality
- $F(3,21) = 3.921$
 $p = 0.023$
- Speech Detector / Human Control
- $F(1,23) = 0.538$
 $p = 0.471$



Experiment Preliminary results

- Dialog Flow Interruptions
- $F(3,21) = 3.208$
 $p = 0.044$

Therapist liked the interruptions



Future improvements

- Responses on long answers
- Allow questions
- Decrease response times
 - Analyze when speaking
 - Faster protocols
 - Sharper timings
- More natural avatar movements
- Anxiety increasing computer responses

Literature

- Parsons, T. D., & Rizzo, A. A. (2008). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, 39(3), 250-261.
- Kessler, R. C., Moguele, K. A., Zhao, S. Y., Nelson, C. B., Hughes, M., Eshleman, S., et al. (1994). Lifetime and 12-Month Prevalence of Dsm-Iii-R Psychiatric-Disorders in the United-States - Results from the National-Comorbidity-Survey. *Archives of General Psychiatry*, 51(1), 8-19
- Hofmann, S. G., & Barlow, D. H. (1999). The costs of anxiety disorders: Implications for psychosocial interventions. *Cost-effectiveness of psychotherapy*, 224-234.
- Klinger, E., Bouchard, S., Legeron, P., Roy, S., Lauer, F., Chemin, I., et al. (2005). Virtual reality therapy versus cognitive behavior therapy for social phobia: A preliminary controlled study. *Cyberpsychology & Behavior*, 8(1), 76-88.
- Brinkman, W.-P., Mast, v. d., 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. *Proceedings of HCI2008 workshop - HCI for technology enhanced learning*, 85-88.



WORKSHOP DISCUSSION

[Willem-Paul Brinkman] You suggest considering for further research the ability of allowing patient to ask the avatar questions. Have to any suggestions how this might be done?

At the moment I have not clear vision on that. But the first step would be the need to detect whether a user response is a question in the first place. Once that has been establish can you start looking at the content of the question.

[Charles van der Mast] A real dialogue between humans is more than an exchange of data or information. How do you support this "more than strings" richness of human communication? Or: how important would it be that the words pronounced by the avatars are spoken by an experienced actor?

All intelligence from the conversation and the sentences are baked in by the human that creates the sentences. The computer is left with the choice between sentences and nothing more. The recording of the sentences is very important because that is all the expressiveness the computer has to offer. So higher quality of the recording could improve the overall feeling of the conversation. It might therefore also be interesting to record one sentence with multiple pronunciations and emotional loading. The editor could handle this and the interpreter program should choose the right pronunciation.