The effect of priming pictures and videos on a question-answer dialog scenario in a virtual environment (Preliminary Version)*

Chao Qu†

Delft University of Technology, 2628CD, Delft, The Netherlands

Willem-Paul Brinkman

Delft University of Technology

Pascal Wiggers

Delft University of Technology

Ingrid Heynderickx

Delft University of Technology & Philips Research Laboratories

October 2, 2013

*Qu, C., Brinkman, W.-P., Wiggers, P., & Heynderickx, I. (2013). The effect of priming pictures and videos on a question-answer dialog scenario in a virtual environment. Presence Teleoperators and Virtual Environments, 22(2), 91-109.

†Correspondence to c.qu@tudelft.nl

Abstract

Having a free speech conversation with avatars in a virtual environment can be desirable in virtual reality applications such as virtual therapy and serious games. However, recognizing and processing free speech seems too ambitious to realize with the current technology. As an alternative, pre-scripted conversations with keyword detection can handle a number of goal-oriented situations as well as some scenarios in which the conversation content is of secondary importance. This is, for example, the case in virtual exposure therapy for the treatment of people with social phobia, where conversation is for exposure and anxiety arousal only. A drawback of pre-scripted dialog is the limited scope of user's answers. The system cannot handle a user's response, which does not match the pre-defined content, other than by providing a default reply. A new method which uses priming material to restrict the possibility of the user's response is proposed in this paper to solve this problem. Two studies were conducted to investigate whether people can be guided to mention specific keywords with video and/or picture primings. Study 1 was a two by two experiment in which participants (n = 20) were asked to answer a number of open questions. Prior to the session, participants watched priming videos or unrelated videos. During the session, they could see priming pictures or unrelated pictures on a whiteboard behind the person who asked the questions. Results showed that participants tended to mention more keywords both with priming videos and pictures. Study 2 shared the same experimental setting but was carried out in virtual reality instead of in the real world. Participants (n = 20) were asked to answer questions of an avatar when they were exposed to priming material before and/or during the conversation session. The same results were found: the surrounding media content had a guidance effect. Furthermore, when priming pictures appeared in the environment, people sometimes forgot to mention the content they typically would mention.

1 Introduction

Virtual reality (VR) is being used increasingly to support cognitive behavior therapy (CBT) especially for exposure exercises. With the advantages of low cost, convenient manipulation and repeatability, virtual reality exposure therapy (VRET) is receiving increasing scientific and public attention (Anderson, Rothbaum, & Hodges, 2001; Anderson, Jacobs, & Rothbaum, 2004; Krijn, Emmelkamp, Olafsson, & Biemond, 2004; Szegedy-Maszak, 2004). The feeling of being immersed, or otherwise stated the feeling of being 'present' in the virtual reality, is an important concept in virtual reality. Without a certain level of presence, the required anxiety level cannot be obtained by the therapy. Presence is the key element to make patients perceive virtual objects, events, entities and environments as if the technology was not involved in the experience (Lombard et al., 2000). A lack of presence is seen as one of the reasons for the relatively high dropout rate for some virtual therapy (Krijn, Emmelkamp, Biemond, et al., 2004).

In VRET for individuals with social phobia, interaction between a patient and a virtual human, i.e. an avatar needs to arouse a certain level of social anxiety (Robillard, Bouchard, Dumoulin, Guitard, & Klinger, 2010). Regulating the response of the avatar automatically to the required realistic level based on the patient's behavior can be hardly realized with current speech processing technology. In current VRET systems, the responses are usually controlled by the therapist, who also needs to monitor the patient in order to deliver the appropriate treatment, which increases the workload of the therapist (Brinkman, Mast, Sandino, Gunawan, & Emmelkamp, 2010). To alleviate this workload, integration of a keyword based dialog manager into a virtual therapy system has been proposed (ter Heijden, Qu, Wiggers, & Brinkman, 2010). Compared to the human control, maintaining the conversation between patient and avatar with keyword detection seems a promising alternative to reduce the workload of the therapist and at the same time, evoke social anxiety at an appropriate level for the patient (ter Heijden & Brinkman, 2011).

In order to use a dialog manager, usually the dialog content needs to be pre-defined. Figure 1 shows an example of a dialog structure. The dark blocks are the computer's responses and the light blocks are the possible types of user response. These types are distinguished by the keywords that appear in the user's response and are linked with the corresponding computer response.

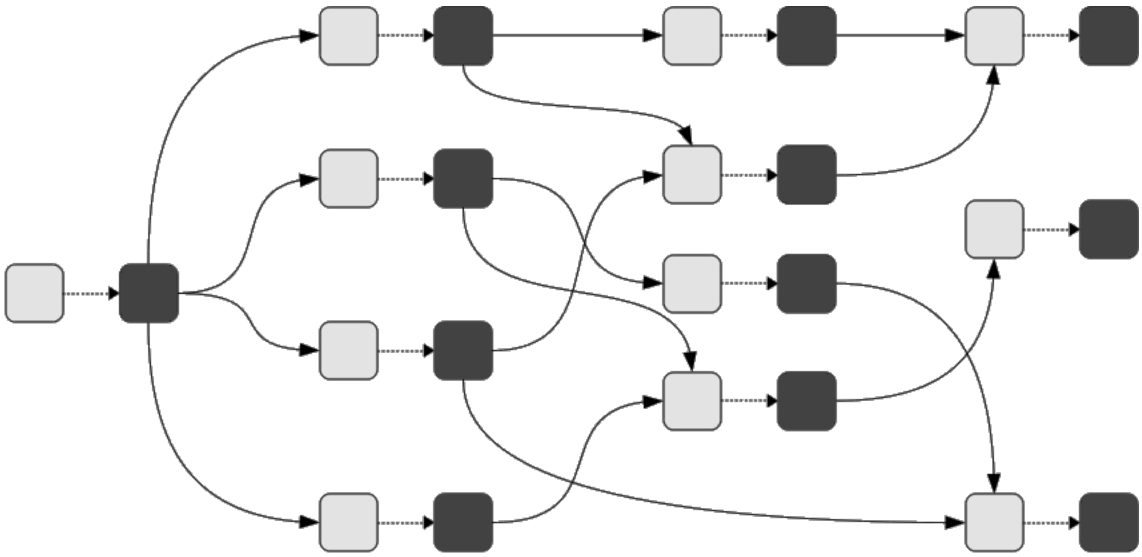


Figure 1: Example of a dialog structure; the dark blocks are the computer's utterances, while the light blocks represent the user's responses. All possible responses are pre-defined.

The main limitation of a pre-scripted dialog manager is that it can only handle a user's response that is in the pre-defined database. Of course, it is possible to define a default response to each question, such as "That's interesting! Tell me more." However, the default response is normally ambiguous. Users may have the feeling that the computer does not really respond to what they are saying. The default response should therefore be only a last remedy. However, if the variety of the user's responses can be restricted to match the set of pre-defined keywords better, the efficiency of the dialog manager could be improved.

Since the computer's response is linked to keywords, two actions can be taken to improve the pre-scripted dialog system: (1) increase the number of keywords in the pre-defined database, or (2) limit the range of responses users are likely to give. Method 1 is a possibility, however an extensive set of keywords is needed. On the other hand, method 2 seems to limit the users' free will, which is not desirable for all applications. For an application such as speech recognition for dictation, method 2 is not desirable. However, for VRET, speech recognition is mainly used to evoke the anxiety patients experience when they are engaged in social interaction.

As a branch of cognitive therapy, VRET inherits the assumption that problematic feelings are different from feelings in general. The problematic feelings are not evoked by reality or certain events, but by the person's cognition about them (Emmelkamp, Bouman, & Scholing, 1992). As long as the conversation is going on, the anxiety provoking stimuli will exist and the system will work well. Therefore it is less important to capture the true meaning of what the patient is saying. Besides, it is less relevant that a person provides an unbiased opinion. So for a VRET system method 2 might be a convenient solution as long as people do not experience that their free will is limited.

In other words, keyword-based speech recognition with a limited set of keywords seems an appropriate technology for evoking anxiety by giving the patient the experience of a social interaction with an avatar, on condition that the patient uses the right keyword. Displaying a list of keywords for the patient to choose from by reading them aloud may result in almost perfect speech recognition, still this might not make the conversation natural (Brinkman, van der Mast, & de Vliegher, 2008). Another approach would be to take advantage of the virtual environment which can be easily controlled. Cues can be integrated into the virtual environment during the conversation between a patient and an avatar, priming the patient to use the pre-defined keywords in his or her answers. For example, if "Paris" is a keyword, there could be a picture of the Eiffel Tower on the wall. With an elaborate virtual environment that includes multiple priming elements, the patient will not have the feeling that his or her free will is limited.

To make this approach successful, people should be influenced sufficiently by these cues to use the expected keywords. The key question therefore is, can priming be used effectively in an VR environment to influence user's responses in a conversation with an avatar?

Two experiments that address this question are described in this paper. The first experiment was conducted in a real-life setting, and focused on the question whether priming is noted in a conversation at all. Having an effect in real life is seen as a pre-condition for extending the study into a VR environment. The second experiment examined whether picture and video priming influenced a user's answers during a conversation with an avatar.

This paper is structured as follows. First, it discusses related work and the theoretical background for social phobia, virtual reality exposure therapy, speech processing, priming, the concept of presence and how it can be measured. Next, the paper introduces the two experiments, including the experimental setting, the procedure and the results. Finally, the results are discussed and some conclusions are given.

2 Theoretical Background

2.1 Social Phobia and Exposure Therapy

Social phobia is one of the most often occurring anxiety disorders: 12.1% of US population (Ruscio et al., 2008), 9.3% of Dutch population (de Graaf, Ten Have, van Gool, & van Dorsselaer, 2012) and in general 6.7% of the European population (Fehm, Pelissolo, Furmark, & Wittchen, 2005) are affected by social phobia during their lifetime. Patients with social phobia suffer from a strong fear of one or more social situations, such as speaking in public, entering a room full of people, shopping, etc. They fear to embarrass themselves in social situations, they feel uncomfortable and try to avoid being exposed to social situations (DSM-IV-TR, 2000).

CBT is often offered as a treatment for social phobia (Fava et al., 2001). Patients are gradually exposed to actual real-life social situations (vivo) or are asked to imagine a social situation (vitro) such as ordering food in a restaurant. Although exposure in vivo, the gold standard, it is an effective treatment, it still has some limitations: the unpredictability of the daily social situation, its dependency on other people in the surrounding (Emmelkamp et al., 2002), and also the effort involved in organizing the social event (Robillard et al., 2010).

2.2 Virtual Reality Exposure Therapy

Virtual Reality technology matured fast in recent decades. The steady increase of computer speed and the improvement of display quality now allow for virtual worlds that are realistic enough to evoke anxiety, though patients are aware that what they see is not real, especially in the situation where they feel phobic (Emmelkamp, Bruynzeel, Drost, & van der Mast, 2001; Walshe, Lewis, O'Sullivan, & Kim, 2005).

Exposing people to virtual reality to treat their phobia is considered as a good alternative to traditional exposure in vivo. Similar to exposure in vivo, patients are subjected to anxiety-provoking stimuli in a gradual order, from the least anxiety provoking stimulus to the most anxiety provoking one. The patients cannot avoid those stimuli and they are allowed to get used to it gradually (Feske & Chambless, 1995; Taylor, 1996; Gould, Buckminster, Pollack, Otto, & Massachusetts, 1997). VRET offers a safer, less costly treatment than exposure in vivo (Klinger et al., 2005; Robillard et al., 2010). It has being studied for treating a number of phobias such as fear of flying (Muhlberger, Wiedemann, & Pauli, 2003; Rothbaum, Hodges, Watson, Kessler, & Opdyke, 1996), fear of height (Krijn, Emmelkamp, Olafsson, & Biemond, 2004; Rothbaum et al., 1995), fear of special insects (Carlin, Hoffman, & Weghorst, 1997; Garcia-Palacios, Hoffman, Carlin, Furness, & Botella, 2002; Botella et al., 2005), and treatment of post-traumatic stress disorder (Difede & Hoffman, 2002). Recent meta-analyses indicate that VRET is as effective as exposure in vivo (Gregg & Tarrier, 2007; Parsons

& Rizzo, 2008; Powers & Emmelkamp, 2008) in treating some phobias such as fear of flying.

Due to the social nature of social phobia, human behavior seems crucial to evoke anxiety. Therefore, compared to VR worlds for other types of phobia, developing a VR world for the treatment of social phobia comes with its own set of challenges such as realistic avatars that face patients. So far, most research focuses on a small set of specific social situations such as speaking in front of a group of avatars (North, North, & Coble, 1998; Slater, Pertaub, & Steed, 1999; Pertaub, Slater, & Barker, 2001; S. R. Harris, Kemmerling, & North, 2002; North, Schoeneman, & Mathis, 2002; Pertaub, Slater, & Barker, 2002; Anderson, Zimand, Hodges, & Rothbaum, 2005; Klinger et al., 2005; Slater, Pertaub, Barker, & Clark, 2006) or ordering food in a restaurant or a bar (James, Lin, Steed, Swapp, & Slater, 2003; Klinger et al., 2004). The variety in avatar behavior is then usually limited to the body posture (Anderson, Rothbaum, & Hodges, 2003; Herbelin, 2005; Klinger et al., 2004; Slater, Antley, et al., 2006) and eye gazing (Riquier, Stankovic, & Chevalley, 2002; Herbelin, Riquier, Vexo, & Thalmann, 2002). Moreover, verbal responses of the avatar are often limited to a small set of pre-recorded responses, or exist of a live voice over by the therapist. A new approach, however, is to use a large set of responses supported by a dialog manager system (ter Heijden & Brinkman, 2011; Brinkman et al., 2012).

2.3 Speech Processing and Dialog Manager

Using speech recognition to analyze what the patient is saying and automatically selecting an appropriate avatar response is a potential way to reduce the workload of the therapist.

Research on free speech conversation between man and machine has a relatively long tradition. An early version of conversation agents are chatbots. A chatbot is a computer program primarily designed for casual conversation (Weizenbaum, 1966; Hutchens & Alder, 1998; Wallace, 2009).

Chatbots simulate an intelligent conversation with one or more human users via auditory or textual methods (Quittner, 1997). The use of sophisticated natural language processing for a chatbot seems ineffective since the speech recognition of oral user input itself is still problematic. The ideal speech recognizer which converts human speech into text words is not existing yet (Jurafsky & Martin, 2000), not to mention free speech processing.

The conversational agents such as real estate agents (Cassell et al., 1999), e-retail (McBreen & Jack, 2001) and automated phone reservation systems (McTear, O'Neill, Hanna, & Liu, 2005) are goal-oriented. They simply scan for keywords within the input and pull a reply with the most matching keywords, or the most similar wording pattern, from a predefined textual database. Other conversational agents like TRINDI (Larsson, 2000) are task-oriented, which means they act on specific information in the dialog context. Although most of these agents have already been put into practical use nowadays, none of them can really understand the real meaning of the casual conversation.

More recent research also focused on patients in virtual reality exposure therapy for social phobia. These studies used automatic keyword detection with semi-scripted dialog controlled by a computer algorithm (ter Heijden et al., 2010; ter Heijden & Brinkman, 2011). The avatars can determine their responses depending on the keywords in the patient's responses. The goal of this approach is to increase a feeling of having an actual free speech conversation, opposite to the situation where the patient reads aloud one of four sentences displayed on a screen (Brinkman et al., 2008). To make the patient's response more predictable, the scenario focuses on specific topics, e.g. a presentation on democracy. However, for these scenarios, there still is a high chance that a patient does not mention any pre-scripted keyword. In that case, the system has to fall back to a default response. In order to avoid this situation, the chance that a patient says certain keywords should be increased without making him or her feel forced or limited during his or her conversation with the avatar.

2.4 Priming Theory

Priming can be seen as the incidental activation of a person's knowledge structure that can lead the person to specific behavior and attitudes (Bargh, Chen, & Burrows, 1996; Bargh, 2006). The use of priming to guide people towards specific verbal responses seems an appropriate mechanism to bias users in favor of giving responses that include specific keywords.

In semantic priming, the prime and the target are from the same semantic category and share features (Ferrand & New, 2003). For example, the word dog is a semantic prime for wolf, because both are similar animals. Semantic priming is theorized to work because of spreading activation in neural circuits in the brain (Reisberg, 2006). When a person thinks of one item in a category, similar items are stimulated by the brain. Even if they are not words, morphemes can also prime for complete words that include them (Marslen-Wilson, Tyler, Waksler, & Older, 1994). An example of this would be that the morpheme 'psych' can prime for the word 'psychology'.

Various studies have examined the concept of priming (Ortells, Vellido, Daza, & Noguera, 2006; Sperber, McCauley, Ragain, & Weil, 1979; Rosch, 1975; Williams & Bargh, 2008; J. L. Harris, Bargh, & Brownell, 2009; Yap, So, Melvin Yap, Tan, & Teoh, 2011), such as daily television advertisement priming (J. L. Harris et al., 2009), masked picture priming with precise time control (Marzouki, Grainger, & Theeuwes, 2007, 2008), colour priming (Mayr, Hauke, Buchner, & Niedeggen, 2009) and temperature priming (Williams & Bargh, 2008). Among these studies, some priming experiments are related to virtual reality (Pena, Hancock, & Merola, 2009; Nunez & Blake, 2003), but most of them explore the theory underlying the priming phenomenon. To our knowledge there are no studies that use priming in the context of supporting question-answer dialogs in virtual reality, or even in reality.

2.5 Presence

The concept of presence contains several very different facets. Generally it covers two sub-concepts: physical (or spatial) presence and social presence (IJsselsteijn, 2000; von der Putten et al., 2012). Physical presence refers to the "sense of being in the virtual environment rather than in the environment in which one is physically located" (Witmer, Jerome, & Singer, 2005). Social presence refers to the feeling of being together with another person (Biocca, Harms, & Gregg, 2001) or the illusion of sharing the same physical space (Riva, Davide, & IJsselsteijn, 2003). This study focused on physical presence since there is no communication between participants and other real humans. Slater (2009) refers to physical presence as 'Place Illusion', which contributes to realistic responses in the virtual environment. A high level of presence would elicit responses in the virtual environment similar to the ones in the real world. If priming in the context of supporting question-answer dialogs works in reality, this should yield a similar effect in a high immersive virtual environment.

Different approaches have been taken to measure presence and generally there are two categories: subjective measurement, i.e. self-reporting during or after the exposure in the virtual environment and objective measurement, i.e. physiological or behavioral response. By far the most common measurement of presence reported in literature is the subjective post-test rating. This type of test is easy and inexpensive to apply, and regarded as an effective approach to measure the concept of presence (IJsselsteijn, 2000; Insko, 2003). Another advantage of a subjective post-test rating is that it does not interfere with the user's experience while in the virtual environment. On the other hand, there are also several limitations to a post-test self-reported measurement. First, it is prone to result into social desirable responses. Participants may guess what the investigator examines, and which outcome he or she expects. They may answer according to or contrary to these predictions (von der Putten et al., 2012). Reliability problems have also been shown (Freeman, Avons, Pearson, & IJsselsteijn, 1999). Second, presence is considered a phenomenon which occurs during the exposure in a virtual environment, a post-experimental test of presence may be more influenced by events towards the end of the immersion. To overcome this issue some researchers use a real-time approach to measure presence (Freeman et al., 1999), e.g. by asking people about their presence experience while being immersed. However, interruptions while being immersed can also affect the presence experience (Hartanto et al., 2012).

Objective measures based on participants' behavioral or physiological responses (e.g. gestures, posture, proxemics, skin conductance, heart rate) can be assessed during the experience of presence. If the participants behave in the virtual world as if they are in an equivalent real world, this means they experience presence. However, a problem with behavioral measures is that there is little likelihood that a behavioral measurement is suitable in all environments (Sanchez-Vives & Slater, 2005). The main problem with physiological measurement is that several different stimuli could produce the same changes in physiological measures (Insko, 2003), and it is not suitable for virtual worlds in which physiological responses are not obvious (Sanchez-Vives & Slater, 2005). Additionally, a pre-measurement is required to offset physiological measurements in the experimental condition, for example with a neutral (stressor free) virtual world (Busscher, de Vliegher, Ling, & Brinkman, 2011).

In the current study, the main focus is on evaluating whether in a virtual world individuals show a similar response pattern to primed and no-primed questions as individuals would do in the real world. This could directly contribute to enhancing human-avatar conversations. To recreate such priming impact, a sufficient level of presence in the virtual world seems a prerequisite. Therefore the study also included an additional subjective presence measurement by asking individuals to complete the post-test subjective Igroup Presence Questionnaire (IPQ) (Schubert, Friedmann, & Regenbrecht, 2001). This questionnaire is widely used (Alsina-Jurnet & Gutierrez-Maldonado, 2010; Freire, Carvalho, Joffily, M., Zin, & Nardi, 2010; Krijn, Emmelkamp, Biemond, et al., 2004; Ling, Brinkman, Nefs, Qu, & Heynderickx, 2013). Therefore, our results can be compared to other studies. The availability of an online IPQ dataset[1] made it possible to examine if at least a similar level of presence was obtained in our experiments as reported in other studies. Failing to do so, would give a probable cause if the priming impact would not be replicated in a virtual world. To keep real world and virtual world conditions similar, the presence measurement was obtained after the exposure to the virtual world, thereby avoiding potential priming interference. As no obvious physiological effects between priming and non-priming conditions were expected, physiological measurements were not regarded as an effective mean to measure presence in this study.

2.6 Hypotheses

In order to test the effectiveness of priming for a question-answer dialog situation, two studies described in the following sections were conducted. Study 1 took place in a real-world setting and Study 2 took place in virtual reality. The two studies aimed at testing the effect of videos and pictures in priming a topic in a limited conversation scenario. Study 1 was a pre-condition of study 2, and seen as a contrast if the priming had no effect in virtual reality.

Pictures and videos were chosen as priming material for two reasons, (1) they were easy to find and commonly seen in daily life, and (2) they could be easily integrated in the virtual environment. As priming material, pictures and videos also played a different role in the experiments. Pictures were used as continuous priming stimulus during the conversation, while the videos were used for upfront priming, as they were only shown before the conversation. The three hypotheses tested in the two experiments were:

H1. (a) priming videos increase the chance that individuals use specific keywords in their answers when having a real-life conversation, (b) priming pictures increase the chance that individuals use specific keywords in their answers when having a real-life conversation.

[1]http://www.igroup.org/pq/ipq/data.php

H2. (a) priming videos increase the chance that users use specific keywords in their answers while having a conversation with an avatar in virtual reality, (b) priming pictures increase the chance that users use specific keywords in their answers while having a conversation with an avatar in virtual reality.

H3. (a) priming videos prevent people to give otherwise common answers, (b) priming pictures prevent people to give otherwise common answers.

Hypothesis 3 considers a potential side-effect of priming, namely that users are less likely to give common answers if they are primed to give a non-common answer.

3 Study 1, Human-Human dialog

Study 1 focused on testing the influence of video and picture priming in a real-world setting, in which two persons had a conversation on a certain topic.

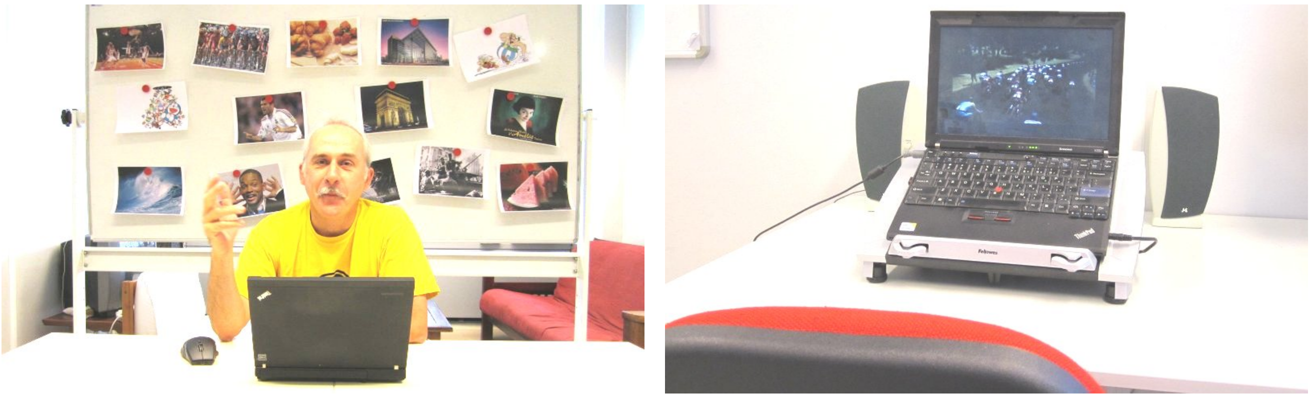
3.1 Experiment Design

The content of the videos and pictures was specially selected for the experiment. The stimuli could be related to the questions of the interviewer, and act as a cue towards specific answers, or the stimuli could be totally unrelated to the topic of the conversation. Therefore, there were two independent variables (video / pictures and related / unrelated), which led to four conditions as can be seen in Table 1.

Table 1: Experiment Conditions

	unrelated picture	**related picture**
unrelated video	condition 1	condition 3
related video	condition 2	condition 4

The experiment had a two-by-two within-subject design and each participant experienced the four conditions. To avoid potential learning effects, four different conversation topics were prepared. The order, in which the four conditions were presented, was counterbalanced in a reduced Latin square (Denes & Keedwell, 1974), while the topics were assigned randomly to the conditions. Participants faced an interviewer to talk about these four topics. Before they had a conversation on a topic, two videos were shown to the participants (in another room, as shown in Figure 2a). During the conversation, pictures were attached on a whiteboard right behind the interviewer, as shown in Figure 2a. Participants were not informed about the priming aspect of the videos or the pictures. Even the interviewer could not see the pictures behind him and he was not informed about the specific keywords either, which made this experiment double-blinded. The pictures were changed by the experimenter without informing the participant who was watching the videos in the other room.



(a) (b)

Figure 2: Priming with Picture/Video in Real World.
(a) with picture (b) with video

In the condition with priming pictures, seven related pictures were placed on a whiteboard together with seven unrelated pictures as a diversion. In the condition with unrelated pictures, all 14 pictures on the whiteboard were unrelated. In the condition with priming videos, two videos were related to a question of the interviewer. In the condition with unrelated videos, the content of the two videos was not related to the topic of the discussion.

3.2 Materials

Four topics, i.e., Democracy, Dogs, France and Penguins, were used. Each topic comprised seven main questions. For each question an answer with specific keyword was identified, which was chosen from a set of possible answers to that question. A picture corresponding to such a keyword was shown in the related picture condition. In the case of the topic Democracy, for example, there was one question "Could you name me some world famous politicians?". The keyword here was "Kennedy". In the related picture condition, a picture of John F. Kennedy was shown on the white board. The related picture was expected to trigger participants to mention that keyword. In the unrelated picture condition, however, a poster of the 3D movie "UP" was shown instead.

In the related video condition, Kennedy's famous "moon landing speech" was shown, while in the unrelated condition, a card trick video was shown. Whereas there was one picture for each question, there were only two priming videos for each topic, corresponding to two questions and consequently two keywords.

Suitable keywords for each question were chosen via a small pilot experiment. All questions were put in the database of a chatbot. 14 pilot participants were asked to have a chat with the chatbot through MSN and their answers were recorded. Based on the frequency by which a keyword was mentioned in the answers, a keywords was selected. Not the most frequent, but the second most frequently mentioned keyword was used. This was done to avoid a potential ceiling effect and to test hypothesis 3. For example, for the question on the famous politician, there was a high chance that participants would mention the current US president Barack Obama with or without being primed. Priming for a less obvious response would make the effect of priming, therefore, more noticeable in the analysis.

3.3 Procedure

The participant was asked to sit in the room with a white board. In the room, the procedure was explained. During this phase, only unrelated pictures were hanging on the white board. After the explanation, the participant was asked to go next door to see two short videos (as shown in Figure 2b). After the participant left the room, the experimenter changed the pictures on the white board.

When the videos were finished, the participant came back to the room with the whiteboard, where the interviewer asked him/her the seven questions related to the topic. The order of these questions was randomly assigned. The answers were recorded. When the conversation was finished, the participant was asked to go to next door's room again to watch new videos. The experimenter quickly changed the pictures on the whiteboard again, after which the participant was questioned about the next topic when he or she was back in the room. This routine continued until all four topics were finished.

At the end of the experiment, a questionnaire was filled out by the participants, asking whether they had noticed that some pictures or videos were related to the topic, and whether these pictures and videos helped them in the conversation.

3.4 Participants

As the dialogs were in Dutch, all selected participants were Dutch speakers (native speakers or people who had at least 5 years of experience in speaking Dutch). Twenty participants took part in the experiment (3 females, 17 males) ranging in age between 25 and 55 years ($M = 28.00, SD = 7.33$). All participants were recruited from Delft University of Technology, 2 were undergraduate students, 10 were master students, 6 were PhD researchers and 2 were university staff. All participants voluntarily took part in this experiment. They only received a small gift (less than 5€) after the experiment.

3.5 Results

There were seven main questions per conversation topic, and each of them had a related keyword. Considering that in the priming conditions, the priming material was shown before and during the entire conversation on a topic, there were no question specific priming elements. In other words, there was a chance for the participants to mention a keyword on one particular question, while the priming was originally meant for another question. In the example of the topic on Democracy, most of the priming pictures hanging on the whiteboard were about politicians. When the participants were asked about the famous politician, they could mention any of the politicians shown on the whiteboard besides John F. Kennedy. In other words, successful priming was achieved when the participants mentioned any of the keywords belonging to that topic regardless if the keyword was originally linked to another question. Therefore the number of targeted keywords, which a participant mentioned throughout the complete conversation on a topic, was used in the analysis. This Keywords Hitting Number per topic (KHN) ranging from zero to seven, was the main measure to test hypothesis 1. In order to examine the effect of priming with videos or pictures, an MANOVA was conducted with KHN as dependent variable and the video and picture conditions (i.e., related vs. unrelated) as two independent within-subject variables. The results showed a significant main effect of priming with pictures ($F(1, 19) = 7.12, p. = 0.015$), and also of priming with videos ($F(1, 19) = 16.47, p. = .001$). No significant two-way interaction between pictures and videos ($F(1, 19) = .05, p. = .728$) was found.

Figure 3 and Table 2 show that on average more keywords were mentioned in the conditions with the priming pictures or videos than in the conditions with unrelated pictures or videos. The result seems to support hypothesis 1. No significant effect was found between the condition with only video priming and the condition with only picture priming ($t(19) = 1.48, p. = .154$). Analysis of the questionnaire indicated that all 20 participants noticed that the videos and pictures were related to the conversation topics. Furthermore, a binomial test found that a significant ($p. = .041$) majority, i.e.

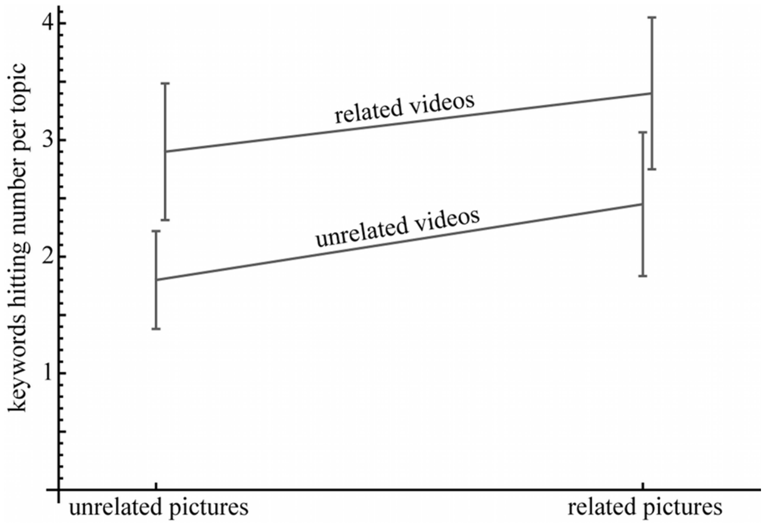


Figure 3: Effect of priming with pictures and videos in the real world based on the mean value of the number of keywords hit per topic and per priming condition, including the 95% confidence interval.

15 of the 20 participants reported the pictures as helpful. This was not found ($p. = .503$) for videos where only 12 participants reported the videos as helpful.

Table 2: Means, Standard Deviations and Bounds in terms of KHN of different conditions in study 1

Condition	Mean	Std. Deviation	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
unrelated picture & unrelated video	1.80	0.89	1.38	2.22
unrelated picture & related video	2.90	1.25	2.31	3.49
related picture & unrelated video	2.45	1.32	1.83	3.07
related picture & related video	3.40	1.39	2.75	4.05

To conclude, the results of study 1 suggest that priming with videos or pictures can result in answers with a specific keyword. The next question was whether it also had a similar effect in a virtual reality environment.

4 Study 2, Human-Avatar dialog

Study 2 was an extension of study 1, aiming at testing the priming influence of the videos and pictures in a virtual environment, while a person had a chat with an avatar on a specific conversation topic. Exactly the same video and picture content, topic questions and experimental setup as in study

1 was used. The interviewer was replicated in an avatar, as shown in Figure 4.

(a) (b)

Figure 4: Experimenter and his Avatar.
(a) Photo (b) Avatar

4.1 Experiment Design

The independent variables were exactly the same as for study 1, i.e. the related/unrelated pictures and videos. The difference was that all pictures and videos were now shown in a virtual reality environment. The pictures were embedded in a virtual picture frame, and the videos were embedded in a virtual television (Figure 6b). The experiment had again a two-by-two within-subject design, with four counterbalanced conditions in a reduced Latin square (Denes & Keedwell, 1974) and four randomly assigned conversation topics, similar to study 1.

4.2 Materials

As mentioned before, topics were Democracy, Dogs, France and Penguins. Questions and keywords were also the same as before. Questions were pre-recorded. The avatars in the virtual environment randomly posed the question and played the recorded question out after the participants pressed the space bar.

4.3 Procedure

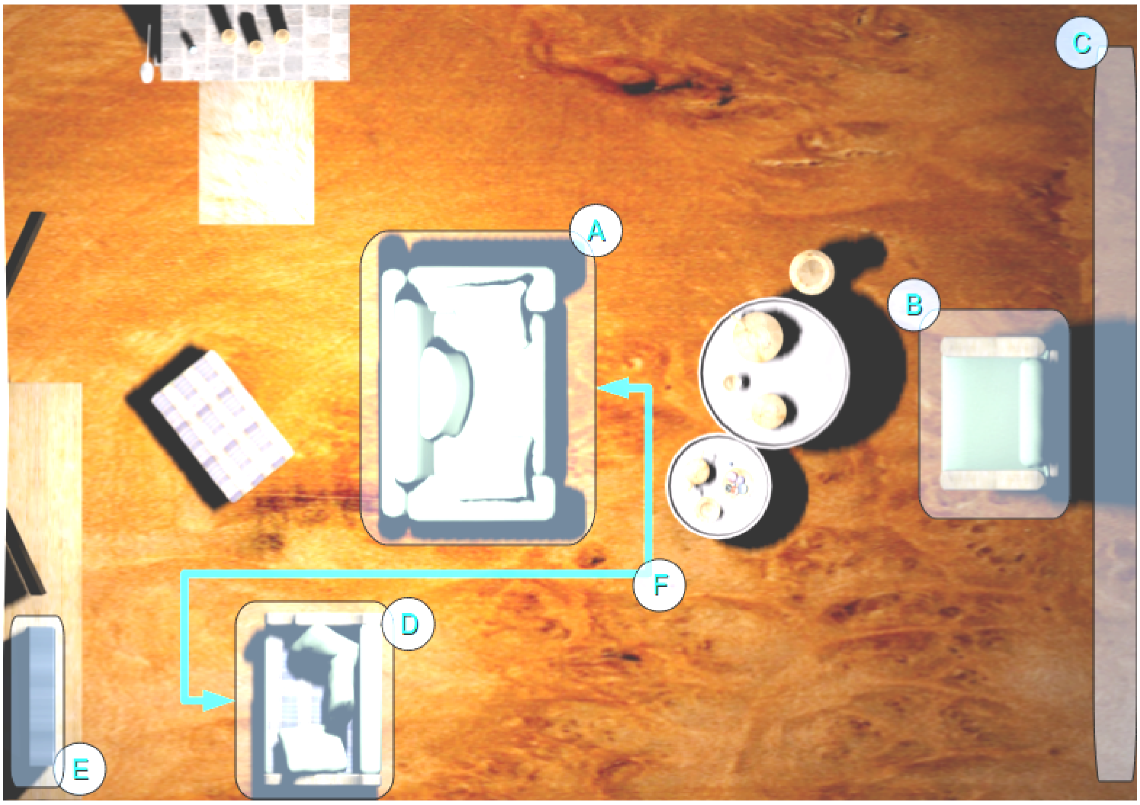


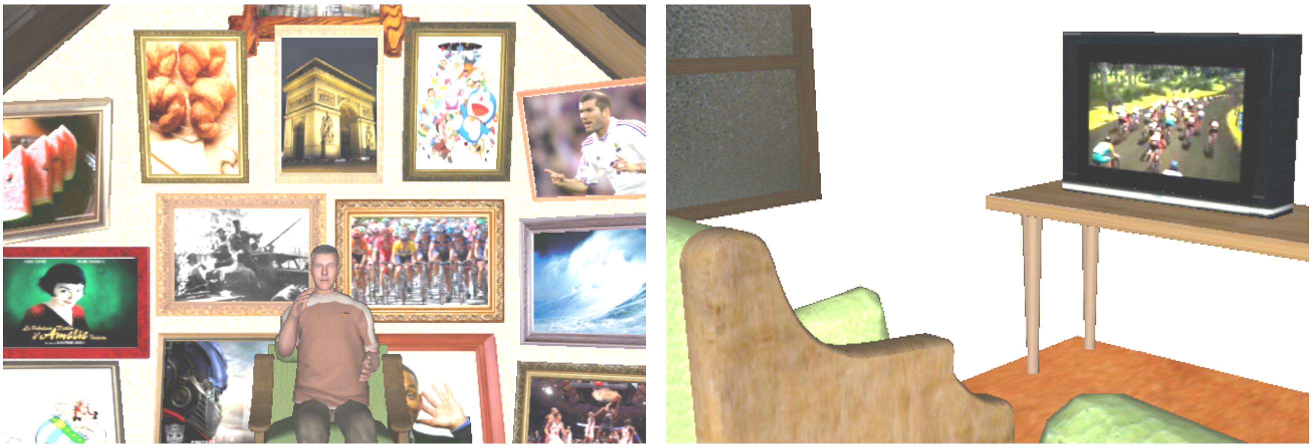
Figure 5: Top View of the Virtual Room for the Experiment.

A small room created with 3Ds MAX was chosen for this experiment, as shown in Figure 5. During the experiment, participants sat on a big sofa in the middle of the room *(A)* while answering the questions of the avatar who was sitting in front of them on another sofa *(B)*. Right behind the avatar was a white wall *(C)* where the pictures were hanging. The participant was able to see the pictures when he or she faced the avatar. The television was put down in a corner of the room *(E)*, with a sofa faced to it on which the participant watched the videos *(D)*. After the participant had seen the two videos, he or she was automatically navigated from sofa *D* to sofa *A* through path *F*. Once the participant was sitting in front of the avatar, the dialog started.

The avatar model was obtained from the Vizard Complete Characters Package[2] and the face part was specially modeled using FaceGen[3]. It was generated based on a three-view photo of the interviewer

[2] `http://www.worldviz.com/products/characters/vcc/index.html`
[3] `http://www.facegen.com`

(Figure 4a, 4b). The questions of the avatar were pre-recorded with the voice of the interviewer from study 1. The 3D models (the room and the avatar) were controlled by Worldviz Vizard[4] 3.0 with programming language Python 2.4.



(a) (b)

Figure 6: Priming with Picture/Video in Virtual World.
(a) with picture (b) with video

Participants wore an eMagin[5] Z800 Head-Mounted Display (HMD) to observe the virtual world. The eMagin Z800 is a USB-powered immersive display device, with a resolution of 800 × 600 pixels. With a build-in 360-degree advanced head tracker, the participant could turn his or her head freely to perceive all the pictures and videos (Figure 6a, 6b). The whole conversation with the avatar was recorded and the participants were asked to complete two questionnaires after they finished all four conditions: (1) the same questionnaire as for study 1, and (2) the Igroup Presence Questionnaire (IPQ) (Schubert et al., 2001). This self-reported presence questionnaire was employed at the end of Study 2 to avoid interfering with the priming effect during the immersion in the virtual environment, and so to create a similar condition in the virtual world.

[4] `http://www.worldviz.com`
[5] `http://www.emagin.com`

4.4 Participants

The participants were again native Dutch speaking people or individuals with at least 5 years of experience in speaking Dutch. Eight female and twelve male participants took part in the experiment. Their age ranged from 18 to 55 years ($M = 27.65$, $SD = 7.64$). All participants were recruited from Delft University of Technology. From the participants, 11 were undergraduate students, 5 were graduated students, 1 was a PhD researcher and 3 were university staffs. None of them participated in study 1. All the participants voluntarily took part in this experiment. They only received a small gift (less than 5€) after the experiment.

4.5 Results

Data Analysis of Study 2

Table 3 shows the result of the Igroup Presence Questionnaire obtained at the end of the experiment, representing the general presence score over all four conditions which the participants experienced. To examine whether the virtual world established a reasonable level of presence, the overall IPQ score was compared to the online IPQ data set[6] for a non-stereoscopic HMD, a procedure also applied in other studies (ter Heijden & Brinkman, 2011; Ling, Brinkman, Nefs, Qu, & Heynderickx, 2012). A MANOVA test was conducted using data source as independent variable and the IPQ general presence and its three subscales (spatial presence, involvement and realism) as dependent variables. No significant difference was found between the online dataset and the IPQ ratings obtained in this experiment ($F(4, 28) = 1.57; p. = .210$), which suggested that participants could immerse themselves at a level that corresponds to presence level reported in other virtual worlds. Assuming this as a reasonable level, the priming material should work in the same way as it had in the real world.

[6]Downloaded on June 9th, 2011. For comparison data see `http://www.igroup.org/pq/ipq/data.php`

Table 3: Means and standard deviations for Igroup Presence Questionnaire

Subscales	Mean	Std. Deviation	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
General Presence (g1)	3.57	1.47	2.94	4.20
Spatial Presence (sp)	3.23	1.30	2.67	3.79
Involvement (inv)	2.82	1.46	2.20	3.45
Realism (real)	2.06	0.99	1.64	2.48

As in study 1, KHN was taken as the main measure. An MANOVA was conducted with KHN as dependent variable and pictures and videos as independent within-subject variables. The results showed a significant main effect for priming with pictures ($F(1, 19) = 13.5, p. = .002$), and for priming with videos ($F(1, 19) = 20.15, p. < .001$). No significant two-way interaction between pictures and videos ($F(1, 19) = 0.33, p. = 0.577$) was found. Again no significant effect was found between the condition with only video priming and the condition with only picture priming ($t(19) = -.33, p. = .748$).

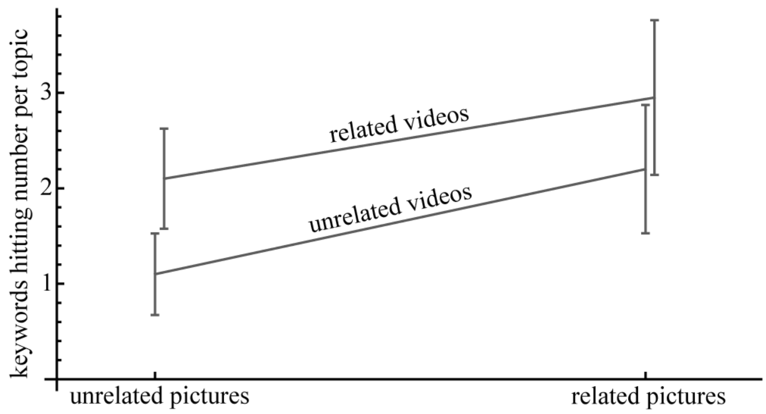


Figure 7: Effect of priming with pictures and videos in the virtual world based on the mean value of the number of keywords hit per topic and per priming condition, including the 95% confidence interval.

Figure 7 and Table 4 show that on average more keywords were mentioned in the conditions with priming pictures or videos than in the conditions with unrelated pictures or videos. This seems to confirm the second hypothesis. Furthermore, the results of the questionnaire indicated that all participants had noticed that the videos and pictures were related to the conversation topics. Again a binomial test found a significant ($p. = .012$) majority, i.e. 16 out of 20 participants reported the pictures as helpful. No significant ($p. = .263$) majority, i.e. 13 out of 20 participants, reported the videos as helpful.

Table 4: Means, Standard Deviations and Bounds in terms of KHN of different conditions in study 2

			95% Confidence Interval	
Condition	Mean	Std. Deviation	Lower Bound	Upper Bound
unrelated picture & unrelated video	1.10	0.91	0.67	1.53
unrelated picture & related video	2.10	1.12	1.58	2.62
related picture & unrelated video	2.20	1.44	1.53	2.87
related picture & related video	2.95	1.73	2.14	3.76

Comparison of Study 1 and Study 2

Both studies had a similar setup, except for the experimental environment, which was the real world in study 1 versus a virtual world in study 2. In order to study the potential effect of the environment, a MANOVA was conducted on the KHN measure, taking environment as a between-subject variable, and video and picture again as two within-subject variables. As expected, the analysis found a significant main effect for picture ($F(1, 38) = 20.57, p. < .001$) and for video ($F(1, 38) = 35.47, p. < .001$). No significant main effect ($F(1, 38) = 3.49, p. = .070$) was found for environment, nor were the two-way interactions between video and environment ($F(1, 38) = 0.22, p. = .641$), and between picture and environment ($F(1, 38) = 1.37, p. = 0.249$) significant. Finally, also the three-way interaction between the independent variables ($F(1, 38) = 0.03, p. = .871$) was not significant.

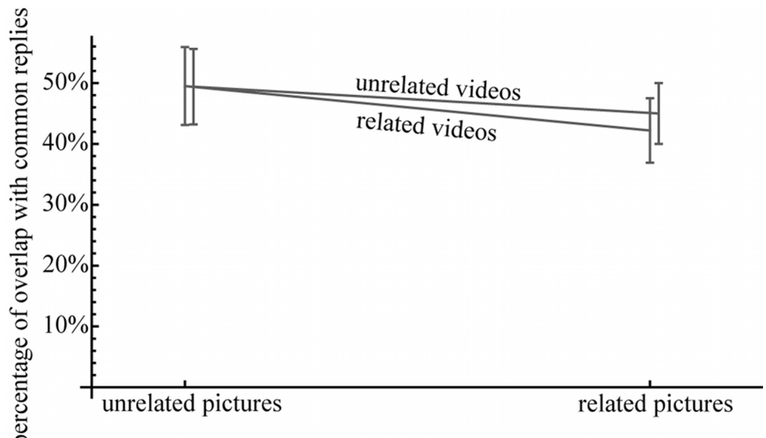


Figure 8: Mean percentage of overlap between the responses given with the top 40% of common responses for each of the priming conditions. The whiskers represent the 95% confidence interval.

The third hypothesis stated that priming prevents people from giving otherwise common answers. The common answer to a question was defined as the answers that were most often given by all participants to that specific question. An arbitrary top of 40% was chosen, excluding the answers primed for, or answers like "don't know". Since for most of the questions, there was more than one common reply, a standardisation was done for each question. For example, in the situation that a question had three common replies, if a participant mentioned two of these common replies in his or her answer, it was counted as 0.66 overlap between the answer given and common replies for this question. An MANOVA test was conducted using the overlap with the top 40% common replies averaged over the seven questions of a topic as dependent variable. The related/unrelated videos and pictures were again taken as independent within-subject variables. Since no difference was found for environment, the data of the two studies were combined in this analysis. The result showed a significant difference for priming with pictures ($F(1, 38) = 5.54, p. = .024$). However, for the video priming content, no such difference was found ($F(1, 38) = 0.22, p. = .640$). The analysis did not show a two-way interaction effect between videos and pictures ($F(1, 38) = 1.12, p. = .296$). Figure 8 and Table 5 show that the percentage of overlap with common replies dropped in the condition with related pictures as compared to the condition with unrelated pictures. This result seems therefore to support only hypothesis 3b.

Table 5: Means, Standard Deviations and Bounds of overlap between the responses given with the top 40% of common responses in different conditions

Condition	Mean	Std. Deviation	95% Confidence Interval	
			Lower Bound	Upper Bound
unrelated picture & unrelated video	0.495	0.197	0.431	0.559
unrelated picture & related video	0.494	0.202	0.431	0.556
related picture & unrelated video	0.422	0.176	0.368	0.475
related picture & related video	0.450	0.170	0.400	0.500

5 Conclusion and Discussion

This paper puts forward three hypotheses regarding increasing and preventing specific answers in a conversation scenario by priming. Hypothesis 1 seems to be confirmed as (a) priming videos and (b) priming pictures increased the chance that the individuals used specific keywords in their answers in a real-life conversation. Similarly, hypothesis 2 was supported for conversations with an avatar in a virtual environment. Important to mention is that no instructions were given to the participants that they should give a specific answer. As all participants noticed that some pictures and videos were related to the topic, it seems likely that some participants might have given, what they thought, socially desirable answers. This, of course, would be their own choice and not compromise their perception of free will. A majority of the participants found the pictures helpful. The majority of the people found the videos helpful, however, this result was not significant. As the videos were only relevant for two of the seven questions within a discussion topic, their objective usefulness was also smaller compared to that of the pictures, which were relevant to all questions.

Priming with videos and pictures in the virtual world seems as effective as in the real world. If there would have been a difference between a virtual world and the real world with a large effect size, e.g. $d = 0.8$, this study would have had a 67% chance to find it (Cohen, 1992). However it was not found. Still the virtual world in this experiment was not an exact copy of the real world. Therefore additional noise could have been created, making the comparison of virtual and real world statistically less powerful. On the other hand, the self-reported presence ratings in this study were comparable to the online IPQ dataset. This result suggests that it was likely that the virtual world did successfully establish enough presence in the participants to evoke a similar priming effect in the virtual world compared to the real world.

With regard to the third hypothesis, only support was found for the effect of the pictures as they were able to prevent individuals from giving otherwise common answers. This suggests that designers of virtual worlds that include conversations, have to consider the potential effect that pictures or objects may have on the user. Even if designers do not plan to use priming material, their objects in the virtual world may already influence what people say to an avatar.

The findings of this study reveal new insights on priming in VR dialogs that can be of practical use to virtual reality exposure therapy or video games. With several pictures in the virtual world or with a short video containing elaborate content, users can be restricted in their conversation up to a certain level, such that it makes the experience of having a conversation with a virtual character more smooth, robust, and hence, natural.

Since there was no significant two-way interaction between videos and pictures, including both videos and pictures as priming material in an application does not seem to have an additional added value. Only using videos as priming material has the advantage that no further manipulation of the VR world is needed during the therapy session or game. Showing several short videos beforehand can already achieve a significant priming effect. However, it might be relatively hard to find a suitable video that primes towards all questions in the conversation. A self-made video might therefore be an interesting alternative. Compared to videos, the advantage of using pictures is that they are much easier to be generated with the appropriate content.

It should also be noticed that the participants were not primed for the most common replies on the questions in the experiment. This was done on purpose to avoid a potential ceiling effect. In an actual application, however, priming the most common reply is more appropriate, and may even result in higher keyword hit rates than the result obtained here.

In the experiment priming was done by hanging the pictures on a wall behind the avatar without any prior knowledge on how much attention these pictures would get. With eye-tracking equipment, however, it is possible to measure where people are actually looking at, and regions of interest may be determined (Redi, Liu, Zunino, & Heynderickx, 2011). It would be interesting to explore to what extent priming can be further enhanced by putting the relevant pictures in the viewer's regions of interest.

As an alternative to pictures, priming can potentially be done by using 3D objects as furniture or decoration. This extension increases the freedom of manipulation in priming elements even further. As a consequence, it is possible to repeat exposure of a patient in the same VR world multiples times. Even if the patient talks about the same topic with the avatar, by changing the priming elements, the content of the conversation can be totally different, exposing the patient to a new experience.

Moreover, it is possible that during an interactive dialogue with a virtual avatar, participants experience a higher level of presence in a primed condition than in an unprimed condition. Since priming may yield a more smooth, natural conversation between the user and the avatar. The user may be less distracted by system limitations, and therefore, may be more easily immersed in the virtual environment. The current experiments did not include interactive communication between the user and the avatars, and so, the effect of priming on presence needs to be studied in future experiments.

Like any empirical study, this study also has a number of limitations. First, the participants did not suffer from social phobia. Therefore, additional research with a group of socially phobic people is needed before making any firm claim about the generalization of our findings to this group. Socially phobic people might be sensitive to socially important cues, such as whether the avatar looks at them, or whether the avatar shows a negative attitude towards them (Clark & McManus, 2002). These negative social interactions are likely to motivate patients not to look directly at the avatar (Chen, Ehlers, Clark, & Mansell, 2002; Horley, Williams, Gonsalvez, & Gordon, 2003; Herbelin et al., 2002; Herbelin, 2005). However, it is not clear whether these patients prefer to gaze at other objects in the environment instead of the avatar. It is also not clear whether gazing towards objects in the environment results in giving more (or less) attention to the priming elements in the room. Moreover, the patients' attention and information processing may be more biased towards specific information because of their higher anxiety level (Amir, 2003), as such reducing the priming effect. To evaluate the viewing behavior of socially phobic people, an eye-tracking device could be used.

Another limitation was that the experimental set-up did not allow a clear comparison between priming before and during the conversation as different stimuli were used, i.e. video versus pictures. Furthermore, the duration of the priming was different for the pictures and the videos. And the videos only focused on two keywords while the pictures focused on all seven keywords.

Despite these shortcomings the results clearly show that priming people beforehand or placing priming material in an environment can increase the number of specific keywords that individuals mention in their communication. This finding opens the door to automatic free speech in VR environments that can be used for the therapy of social phobia patients.

Acknowledgements

This work is supported in part by the Chinese Scholarship Council (No.2008609199) and the COMMIT project - Interaction for Universal Access.

References

Alsina-Jurnet, I., & Gutierrez-Maldonado, J. (2010). Influence of personality and individual abilities on the sense of presence experienced in anxiety triggering virtual environments. *International Journal of Human-Computer Studies*, *68*(10), 788–801.

Amir, N. (2003, November). Attentional bias to threat in social phobia: facilitated processing of threat or difficulty disengaging attention from threat? *Behaviour Research and Therapy*, *41*(11), 1325–1335.

Anderson, P. L., Jacobs, C., & Rothbaum, B. O. (2004). Computer-supported cognitive behavioral treatment of anxiety disorders. *Journal of Clinical Psychology*, *60*(3), 253–267.

Anderson, P. L., Rothbaum, B. O., & Hodges, L. F. (2001). Virtual reality: using the virtual world to improve quality of life in the real world. *Bulletin of the Menninger Clinic*, *65*(1), 78–91.

Anderson, P. L., Rothbaum, B. O., & Hodges, L. F. (2003). Virtual Reality Exposure in the Treatment of Social Anxiety. *Cognitive And Behavioral Practice*, *10*(3), 240–47.

Anderson, P. L., Zimand, E., Hodges, L. F., & Rothbaum, B. O. (2005, January). Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and anxiety*, *22*(3), 156–8. doi: 10.1002/da.20090

Bargh, J. A. (2006, January). What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European journal of social psychology*, *36*(2), 147–168.

Bargh, J. A., Chen, M., & Burrows, L. (1996, August). Automaticity of social behavior: direct effects of trait construct and stereotype-activation on action. *Journal of personality and social psychology*, *71*(2), 230–44.

Biocca, F., Harms, C., & Gregg, J. (2001). The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. *4th Annual International Workshop on Presence*.

Botella, C. M., Juan, M. C., Banos, R. M., Alcaniz, M., Guillen, V., & Rey, B. (2005, April). Mixing realities? An application of augmented reality for the treatment of cockroach phobia. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, *8*(2), 162–71.

Brinkman, W.-P., Hartanto, D., Kang, N., de Vliegher, D., Kampmann, I. L., Morina, N., ... Neerincx, M. A. (2012). A virtual reality dialogue system for the treatment of social phobia. In *Chi'12 extended abstracts on human factors in computing systems* (pp. 1099–1102).

Brinkman, W.-P., Mast, C. V. D., Sandino, G., Gunawan, L. T., & Emmelkamp, P. M. G. (2010, July). 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.

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. *Proceedings of HCI2008 Workshop HCI*, 83–95.

Busscher, B., de Vliegher, D., Ling, Y., & Brinkman, W.-P. (2011). Physiological measures and self-report to evaluate neutral virtual reality worlds. *Journal of CyberTherapy and Rehabilitation*, *4*(1), 15–25.

Carlin, A. S., Hoffman, H. G., & Weghorst, S. (1997). Virtual reality and tactile augmentation in the treatment of spider phobia: a case report. *Behaviour research and therapy*, *35*(2), 153–158.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson, H., & Yan, H. (1999). Embodiment in conversational interfaces. In *Proceedings of the sigchi conference on human factors in computing systems the chi is the limit chi 99* (pp. 520–527). ACM Press.

Chen, Y. P., Ehlers, A., Clark, D. M., & Mansell, W. (2002). Patients with generalized social phobia direct their attention away from faces. *Behaviour Research and Therapy*, *40*(6), 677–687.

Clark, D. M., & McManus, F. (2002, January). Information processing in social phobia. *Biological psychiatry*, *51*(1), 92–100.

Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, *112*(1), 155–159.

de Graaf, R., Ten Have, M., van Gool, C., & van Dorsselaer, S. (2012). Prevalence of mental disorders, and trends from 1996 to 2009. Results from NEMESIS-2. *Tijdschr Psychiatr*, *54*(1), 27–38.

Denes, J., & Keedwell, A. D. (1974). *Latin squares and their applications.* New York: Academic Press.

Difede, J., & Hoffman, H. G. (2002, December). Virtual reality exposure therapy for World Trade Center Post-traumatic Stress Disorder: a case report. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, *5*(6), 529–35.

DSM-IV-TR. (2000). Diagnostic and statistical magnual of mental disorders (Revised 4th ed.). *American Psychiatric Association.*

Emmelkamp, P. M. G., Bouman, T. K., & Scholing, A. (1992). *Anxiety Disorders: A Practitioner's Guide* (1st ed.). John Wiley & Sons.

Emmelkamp, P. M. G., Bruynzeel, M., Drost, L., & van der Mast, C. A. P. G. (2001, June). Virtual reality treatment in acrophobia: a comparison with exposure in vivo. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, *4*(3), 335–9.

Emmelkamp, P. M. G., Krijn, M., Hulsbosch, a. M., de Vries, S., Schuemie, M. J., & van der Mast, C. A. P. G. (2002, May). Virtual reality treatment versus exposure in vivo: a comparative evaluation in acrophobia. *Behaviour research and therapy*, *40*(5), 509–16.

Fava, G. A., Grandi, S., Rafanelli, C., Ruini, C., Conti, S., & Belluardo, P. (2001). Long-term outcome of social phobia treated by exposure. *Psychological Medicine*, *31*(5), 899–905.

Fehm, L., Pelissolo, A., Furmark, T., & Wittchen, H.-U. (2005, August). Size and burden of social phobia in Europe. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology*, *15*(4), 453–62.

Ferrand, L., & New, B. (2003). Semantic and associative priming in the mental lexicon. *Mental lexicon: Some words to talk about words*, 25–43.

Feske, U., & Chambless, D. (1995, January). Cognitive behavioral versus exposure only treatment for social phobia: A meta-analysis. *Behavior Therapy*, *26*(4), 695–720.

Freeman, J., Avons, S. E., Pearson, D. E., & IJsselsteijn, W. A. (1999). Effects of sensory information and prior experience on direct subjective ratings of presence. *Presence-Teleoperators and Virtual Environments*, *8*(1), 1–13.

Freire, R. C., Carvalho, M. R. D., Joffily, M., Zin, W. A., & Nardi, A. E. (2010). Anxiogenic properties of a computer simulation for panic disorder with agoraphobia. *Journal of Affective Disorders*, *125*(1-3), 301–306.

Garcia-Palacios, A., Hoffman, H. G., Carlin, A. S., Furness, T. a., & Botella, C. M. (2002, September). Virtual reality in the treatment of spider phobia: a controlled study. *Behaviour research and therapy*, *40*(9), 983–93.

Gould, R., Buckminster, S., Pollack, M., Otto, M., & Massachusetts, L. (1997). Cognitive-Behavioral and Pharmacological Treatment for Social Phobia: A Meta-Analysis. *Clinical Psychology: Science and Practice*, *4*(4), 291–306.

Gregg, L., & Tarrier, N. (2007, May). Virtual reality in mental health : a review of the literature. *Social psychiatry and psychiatric epidemiology*, *42*(5), 343–54.

Harris, J. L., Bargh, J. A., & Brownell, K. D. (2009, July). Priming effects of television food advertising on eating behavior. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*, *28*(4), 404–13.

Harris, S. R., Kemmerling, R. L., & North, M. M. (2002, December). Brief virtual reality therapy for public speaking anxiety. *Cyberpsychology & behavior*, *5*(6), 543–550.

Hartanto, D., Kang, N., Brinkman, W.-P., Kampmann, I. L., Morina, N., Emmelkamp, P. M. G., & Neerincx, M. A. (2012). Automatic mechanisms for measuring subjective unit of discomfort. *Annual Review of Cybertherapy and Telemedicine*, *181*, 192–196. doi: 10.3233/978-1-61499-121-2-192

Herbelin, B. (2005). *Virtual reality exposure therapy for social phobia* (Vol. 3351). Unpublished doctoral dissertation, Louis Pasteur University.

Herbelin, B., Riquier, F., Vexo, F., & Thalmann, D. (2002). Virtual reality in cognitive behavioral therapy: a study on social anxiety disorder. In *8th international conference on virtual systems and multimedia, vsmm02* (pp. 1–10).

Horley, K., Williams, L. M., Gonsalvez, C., & Gordon, E. (2003, January). Social phobics do not see eye to eye: a visual scanpath study of emotional expression processing. *Journal of Anxiety Disorders*, *17*(1), 33–44.

Hutchens, J. L., & Alder, M. D. (1998). Introducing MegaHAL. *In nemlap3, Conll98 workshop on human-computer conversation, ACL*, 271–274.

IJsselsteijn, W. A. (2000). Presence: concept, determinants, and measurement. *Proceedings of SPIE*, *31*(0), 520–529. doi: 10.1117/12.387188

Insko, B. (2003). Measuring presence: Subjective, behavioral and physiological methods. *EMERGING COMMUNICATION*.

James, L. K., Lin, C.-Y., Steed, A., Swapp, D., & Slater, M. (2003, June). Social anxiety in virtual environments: results of a pilot study. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, *6*(3), 237–43.

Jurafsky, D., & Martin, J. H. (2000). *Speech and Language Processing.* Prentice Hall.

Klinger, E., Bouchard, S., Legeron, P., Roy, S., Lauer, F., Chemin, I., & Nugues, P. (2005). Virtual reality therapy versus cognitive behavior therapy for social phobia: A preliminary controlled study. *Cyberpsychology & behavior*, *8*(1), 76–88.

Klinger, E., Legeron, P., Roy, S., Chemin, I., Lauer, F., & Nugues, P. (2004, January). Virtual Reality Exposure in the Treatment of Social Phobia. *Studies in health technology and informatics*, *99*, 91–119.

Krijn, M., Emmelkamp, P. M. G., Biemond, R., de Wilde de Ligny, C., Schuemie, M. J., van der Mast, C. A. P. G., & de Ligny, C. D. (2004, February). Treatment of acrophobia in virtual reality: The role of immersion and presence. *Behaviour Research and Therapy*, *42*(2), 229–239.

Krijn, M., Emmelkamp, P. M. G., Olafsson, R. P., & Biemond, R. (2004, July). Virtual reality exposure therapy of anxiety disorders: a review. *Clinical psychology review*, *24*(3), 259–81.

Larsson, S. (2000, September). Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural language engineering*, *6*(3&4), 323–340.

Ling, Y., Brinkman, W.-P., Nefs, H. T., Qu, C., & Heynderickx, I. (2012). Effects of Stereoscopic Viewing on Presence, Anxiety and Cybersickness in a Virtual Reality Environment for Public Speaking. *Presence: Teleoperators and Virtual Environments*, *21*(3), 254–267.

Ling, Y., Brinkman, W.-P., Nefs, H. T., Qu, C., & Heynderickx, I. (2013). The relationship between individual characteristics and experienced presence (accepted). *Computers in Human Behavior*.

Lombard, M., Ditton, T., Crane, D., Davis, B., Gil-Egui, G., Horvath, K., . . . Park, S. (2000). Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument. In *Third international workshop on presence, delft, the netherlands.*

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, *101*(1), 3.

Marzouki, Y., Grainger, J., & Theeuwes, J. (2007, September). Exogenous spatial cueing modulates subliminal masked priming. *Acta psychologica*, *126*(1), 34–45.

Marzouki, Y., Grainger, J., & Theeuwes, J. (2008, September). Inhibition of return in subliminal letter priming. *Acta psychologica*, *129*(1), 112–20.

Mayr, S., Hauke, R., Buchner, A., & Niedeggen, M. (2009, April). No evidence for a cue mismatch in negative priming. *Quarterly journal of experimental psychology (2006)*, *62*(4), 645–52.

McBreen, H. M., & Jack, M. A. (2001). Evaluating humanoid synthetic agents in e-retail applications. *IEEE Transactions on Systems Man and Cybernetics Part A*, *31*(5), 394–405.

McTear, M. F., O'Neill, I., Hanna, P., & Liu, X. (2005, March). Handling errors and determining confirmation strategies - An object-based approach. *Speech Communication*, *45*(3), 249–269.

Muhlberger, A., Wiedemann, G. C., & Pauli, P. (2003). Efficacy of a one-session virtual reality exposure treatment for fear of flying. *Psychotherapy Research*, *13*(3), 323–336.

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

North, M. M., Schoeneman, C. M., & Mathis, J. R. (2002). Virtual Reality Therapy: case study of fear of public speaking. *Studies In Health Technology And Informatics*, *85*, 318–320.

Nunez, D., & Blake, E. (2003). Conceptual priming as a determinant of presence in virtual environments. In *Proceedings of the 2nd international conference on computer graphics, virtual reality, visualisation and interaction in africa - afrigraph '03* (pp. 101–108). New York, New York, USA: ACM Press.

Ortells, J. J., Vellido, C., Daza, M. T., & Noguera, C. (2006). Semantic priming effects with and without perceptual awareness. *Psicológica*, *27*, 225–242.

Parsons, T. D., & Rizzo, A. A. (2008, September). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis. *Journal of behavior therapy and experimental psychiatry*, *39*(3), 250–61.

Pena, J., Hancock, J. T., & Merola, N. a. (2009, September). The Priming Effects of Avatars in Virtual Settings. *Communication Research*, *36*(6), 838–856.

Pertaub, D.-P., Slater, M., & Barker, C. (2001, January). An experiment on fear of public speaking in virtual reality. *Studies in health technology and informatics*, *81*, 372–8.

Pertaub, D.-P., Slater, M., & Barker, C. (2002). An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators & Virtual Environments*, *11*(1), 68–78.

Powers, M. B., & Emmelkamp, P. M. G. (2008, January). Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of anxiety disorders*, *22*(3), 561–9.

Quittner, J. (1997). Techwatch: What's hot in Bots. *Time Magazine*.

Redi, J., Liu, H., Zunino, R., & Heynderickx, I. (2011). Interactions of visual attention and quality perception. In *Society of photo-optical instrumentation engineers (spie) conference series* (Vol. 7865, p. 26).

Reisberg, D. (2006). *Cognition: Exploring the science of the mind.* WW Norton.

Riquier, F., Stankovic, M., & Chevalley, A. F. (2002). Virtual gazes for social exposure: Margot and Snow White. In *Proceedings of the 1st. international workshop on virtual reality rehabilitation.*

Riva, G., Davide, F., & IJsselsteijn, W. A. (2003). *Being There: Concepts, Effects and Measurement of User Presence in Synthetic Environments.* Ios Press, Amsterdam.

Robillard, G., Bouchard, S., Dumoulin, S., Guitard, T., & Klinger, E. (2010). Using virtual humans to alleviate social anxiety: preliminary report from a comparative outcome study. *Studies In Health Technology And Informatics*, *154*, 57–60.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*(3), 192–233.

Rothbaum, B. O., Hodges, L. F., Kooper, R., Opdyke, D., Williford, J., & North, M. M. (1995, January). Virtual reality graded exposure in the treatment of acrophobia: A case report. *Behavior Therapy*, *26*(3), 547–554.

Rothbaum, B. O., Hodges, L. F., Watson, B., Kessler, G., & Opdyke, D. (1996). Virtual reality exposure therapy in the treatment of fear of flying: A case report. *Behaviour Research and Therapy*, *34*(5-6), 477–481.

Ruscio, a. M., Brown, T. a., Chiu, W. T., Sareen, J., Stein, M. B., & Kessler, R. C. (2008, January). Social fears and social phobia in the USA: results from the National Comorbidity Survey Replication. *Psychological medicine*, *38*(1), 15–28.

Sanchez-Vives, M. V., & Slater, M. (2005, April). From presence to consciousness through virtual reality. *Nature reviews. Neuroscience*, *6*(4), 332–9. doi: 10.1038/nrn1651

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The Experience of Presence: Factor Analytic Insights. *Presence Teleoperators Virtual Environments*, *10*(3), 266–281.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences*, *364*(1535), 3549–3557.

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., . . . Sanchez-Vives, M. V. (2006, January). A virtual reprise of the Stanley Milgram obedience experiments. *PloS one*, *1*(1), e39.

Slater, M., Pertaub, D.-P., Barker, C., & Clark, D. M. (2006, October). An experimental study on fear of public speaking using a virtual environment. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, *9*(5), 627–33.

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.

Sperber, R., McCauley, C., Ragain, R., & Weil, C. (1979). Semantic priming effects on picture and word processing. *Memory & Cognition*, *7*(5), 339–345.

Szegedy-Maszak, M. (2004). Conquering our phobias: the biological underpinnings of paralyzing fears. *US news world report*, *137*(20), 66–72, 74.

Taylor, S. (1996). Meta-analysis of cognitive-behavioral treatments for social phobia. *Journal of Behavior Therapy and Experimental Psychiatry*, *27*(1), 1–9.

ter Heijden, N., & Brinkman, W.-P. (2011). Design and Evaluation of a Virtual Reality Exposure Therapy System with Automatic free Speech Interaction. *Journal of CyberTherapy & Rehabilitation*, *4*(1), 35–49.

ter Heijden, N., Qu, C., Wiggers, P., & Brinkman, W.-P. (2010). Developing a Dialogue Editor to Script Interaction between Virtual Characters and Social Phobic Patient. In *Proceedings of the ecce2010 workshop - cognitive engineering for technology in mental health care and rehabilitation* (pp. 978–94).

von der Putten, A. M., Klatt, J., Ten Broeke, S., McCall, R., Kramer, N. C., Wetzel, R., . . . Oppermann, L. (2012). Subjective and behavioral presence measurement and interactivity in the collaborative augmented reality game TimeWarp. *Interacting with computers*, *24*(4), 317–325.

Wallace, R. S. (2009). The Anatomy of A.L.I.C.E. *Parsing the Turing Test*, *Part III*, 181–210.

Walshe, D., Lewis, E., O'Sullivan, K., & Kim, S. I. (2005, December). Virtually driving: are the driving environments "real enough" for exposure therapy with accident victims? An explorative study. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, *8*(6), 532–7. doi: 10.1089/cpb.2005.8.532

Weizenbaum, J. (1966). ELIZA - a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45.

Williams, L. E., & Bargh, J. A. (2008, October). Experiencing physical warmth promotes interpersonal warmth. *Science (New York, N.Y.)*, *322*(5901), 606–7.

Witmer, B. G. B. G., Jerome, C. C. J., & Singer, M. J. (2005, June). The factor structure of the presence questionnaire. *Presence-Teleoperators and Virtual Environments*, *14*(3), 298–312. doi: 10.1162/105474605323384654

Yap, D.-F., So, W.-C., Melvin Yap, J.-M., Tan, Y.-Q., & Teoh, R.-L. S. (2011, January). Iconic gestures prime words. *Cognitive science*, *35*(1), 171–83.

Preliminary version of: **Qu, C., Brinkman, W.-P., Wiggers, P., & Heynderickx, I. (2013). The effect of priming pictures and videos on a question-answer dialog scenario in a virtual environment.** *Presence Teleoperators and Virtual Environments*, **22(2), 91-109.**

Figure 1: Example of a dialog structure; the dark blocks are the computer's utterances, while the light blocks represent the user's responses. All possible responses are pre-defined. (one column width)

Figure 2: Priming with Picture/Video in Real World.

Figure 2A: with picture (one column width)

Figure 2B: with video (one column width)

Figure 3: Effect of priming with pictures and videos in the real world based on the mean value of the number of keywords hit per topic and per priming condition, including the 95% confidence interval. (one column width)

Figure 4: Experimenter and his Avatar (one column width)

Figure 4A: Photo

Figure 4B: Avatar

Figure 5: Top View of the Virtual Room for the Experiment (one column width)

Figure 6: Priming with Picture/Video in Virtual World

Figure 6A: with picture (one column width)

Figure 6B: with video (one column width)

Figure 7: Effect of priming with pictures and videos in the virtual world based on the mean value of the number of keywords hit per topic and per priming condition, including the 95% confidence interval. (one column width)

Figure 8: Mean percentage of overlap between the responses given with the top 40% of common responses for each of the priming conditions. The whiskers represent the 95% confidence interval. (one column width)