Talking with a Virtual Human: Controlling the Human Experience and Behavior in a Virtual Conversation

Talking with a Virtual Human: Controlling the Human Experience and Behavior in a Virtual Conversation

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universtiteit Delft,
op gezag van de Rector Magnificus Prof. ir. K.C.A.M. Luyben,
voorzitter van het College van Promoties,
in het openbaar te verdedigen op 3 september 2014 om 15:00 uur
door

Chao QU

M.Sc. in Physical Electronics from Southeast University, China,
Bachelor in Electrical Engineering from Northeast Normal University, China,
born in Suzhou, China.

Dit proefschrift is goedgekeurd door de promotor:

Prof.dr. I.E.J. Heynderickx

Toegevoegd promotor:

Dr.ir. W.P. Brinkman

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. I.E.J. Heynderickx,	Technische Universiteit Eindhoven, promotor
Dr. ir. W.P. Brinkman,	Technische Universiteit Delft, copromotor
Prof. dr. E. Eisemann ,	Technische Universiteit Delft
Prof. dr. H. de Ridder,	Technische Universiteit Delft
Prof. dr. W.A. IJsselsteijn,	Technische Universiteit Eindhoven
Prof. dr. D.K.J. Heylen,	University of Twente
Prof. dr. M. Alcañiz,	Polytechnic University of Valencia
Prof. dr. M.A. Neerincx,	Technische Universiteit Delft (reservelid)

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

Keywords: virtual reality, virtual human, human behavior, emotion, human computer interaction

ISBN 978-94-6203-634-5

Printed by CPI-Koninklijke Wöhrmann Print Service, the Netherlands.

To my beloved parents and wife.

Summary

Virtual humans are often designed to replace real humans in virtual reality applications for e.g., psychotherapy, education and entertainment. In general, applications with virtual humans are created for modifying a person's knowledge, beliefs, attitudes, emotions or behaviors. Reaching these intended goals, however, strongly depends on being able to control the conversation in these applications. Obviously important aspects to control such a conversation are speech recognition and natural language understanding and generation, but besides these aspects also the behavior of virtual humans and objects in the virtual environment may potentially influence the simulated conversation, and therefore, its effectiveness. Understanding which factors in a virtual environment may affect the dialog between a human and a virtual human, and finding ways to control the human experience and behavior during the conversation are the main aims of this thesis.

Three main elements that characterize a conversation between a human and a virtual human were identified, i.e., the surrounding environment, the virtual conversation partner, and the virtual bystanders. Four separated empirical studies were conducted to investigate the effect of these three main elements in the domain of virtual reality exposure therapy for treating social anxiety disorders. The results show that priming materials in the virtual environment such as videos and pictures have a guiding effect on humans having a conversation with a virtual human. Also, emotions expressed when the virtual human speaks are perceived as more intense than emotions expressed when the virtual human listens, and emotions expressed while speaking had a larger effect on people's valence and discussion satisfaction. Furthermore, a positive attitude of the virtual conversation partner, i.e., a happy facial expression while constantly looking at the human conversation partner, and speaking with a positive voice intonation, elicits a more positive emotional state in humans as compared to a negative attitude, i.e., an angry facial expression while looking at the human conversation partner, and speaking with a negative voice intonation. Similarly, a positive attitude of virtual bystanders towards a person, i.e., happy facial expressions and whispering positive comments about the person's behavior, evokes more self-efficacy and less anxiety showing less avoidance behavior in the person compared to a negative attitude of the bystanders, i.e., angry facial expressions and whispering negative comments.

In conclusion, by manipulating virtual objects, the virtual conversation partner or virtual bystanders, a therapist may affect the behavior, emotions and beliefs of a person.

Samenvatting

Virtuele mensen zijn vaak ontworpen om echte mensen in virtual-reality toepassingen, zoals psychotherapie, onderwijs en vermaak, te vervangen. Over het algemeen worden virtuele mensen gemaakt om iemands kennis, aannames, houding, emotie of gedrag te veranderen. Het bereiken van deze gestelde doelen hangt echter sterk af van de mogelijkheid een gesprek te sturen. Belangrijke aspecten om zo'n gesprek te sturen zijn vanzelfsprekend spraakherkenning en begrip van natuurlijk taalbegrip en generatie, maar behalve deze aspecten kunnen ook het gedrag van virtuele mensen en objecten in de virtuele omgeving het gesimuleerde gesprek potentieel benvloeden, en daarmee de effectiviteit. Het begrijpen van welke factoren in de virtuele omgeving de dialoog tussen mens en virtuele mens benvloeden en het vinden van manieren om menselijke ervaring en gedrag te benvloeden tijdens een gesprek zijn de hoofddoelen van dit proefschrift.

Drie hoofdelementen die een gesprek tussen mens en virtuele mens karakteriseren werden gedentificeerd, namelijk de omgeving, de virtuele gesprekspartner en de virtuele omstanders. Vier onafhankelijke empirische studies zijn gedaan om het effect van deze drie hoofdelementen in het domein van Virtual Reality Exposure Therapie voor de behandeling van sociale-angst stoornissen te onderzoeken. De resultaten laten zien dat een voorvertoning van materialen in de virtuele omgeving zoals video's en afbeeldingen een leidend effect heeft op de conversatie met een virtuele mens. Emoties worden ook als sterker ervaren wanneer een virtuele mens spreekt, dan wanneer een virtuele mens luistert, en de geuite emoties tijdens het spreken hadden een groter effect op de positieve of negatieve emotionele toestand en tevredenheid over de discussie. Een positieve houding van de virtuele gesprekspartner, dat wil zeggen een gelukkige gelaatsuitdrukking en spreken met een positieve stemintonatie, wekt bovendien een meer positieve emotionele toestand op in mensen, in vergelijking met een negatieve houding, dat wil zeggen een boze gelaatsuitdrukking bij het aankijken van de gesprekspartner en spreken met een negatieve stemintonatie. Zo leidt ook een positieve houding van de virtuele omstanders ten aanzien van de persoon, dat wil zeggen een blije gelaatsuitdrukking en gefluisterde positieve opmerkingen over de taakuitvoering van de persoon, tot meer zelfovertuiging over de eigen bekwaamheid in de specifieke taak en tot minder angst door ook minder vermijdingsgedrag te vertonen dan bij een negatieve houding van de omstanders, dat wil zeggen een boze gelaatsuitdrukking en gefluisterde negatieve opmerkingen.

Concluderend kan gezegd worden dat een therapeut het gedrag, de emoties en aannames van een persoon kan benvloeden door virtuele objecten, virtuele gesprekspartners of de virtuele omstanders te manipuleren.

Contents

Chapter 1

Introduction

Virtual humans are computer-generated characters that exist of a visual body with a humanlike appearance and may express a range of observable behaviour. They are often designed to replace actual humans in virtual environments for e.g., entertainment, education, and psychotherapy. More specifically, virtual humans can provide a human-like interface to information services (Vandeventer and Barbour, 2010), act as a museum guide (Foutz et al., 2012; Kopp et al., 2005), play characters in entertainment systems (Balcisoy et al., 2000; Dow et al., 2007; Mateas and Stern, 2003), or act as a role player in training systems such as clinical interviews (Kenny et al., 2008), public speaking (Slater et al., 1999), sales conversations (Muller et al., 2012), negotiation conversations (Broekens et al., 2012; Core et al., 2006; Traum et al., 2003), or an army mission rehearsal system for teaching critical decision-making skills (Hill et al., 2003). Ideally these virtual conversations are conducted through natural language speech, but in practice synthetic speech was regularly implemented. Virtual conversations in general have an intended purpose or goal, being the modification of a person's knowledge, beliefs, attitudes, emotions or behaviour. The ability to control the conversation has a direct impact on the ability to meet this intended goal. Besides aspects as speech and language processing and generation, the behaviour of objects and characters in a virtual environment may potentially influence the simulated conversation, and therefore, its effectiveness. Understanding which factors in a virtual environment affect the dialog between a human and virtual human, and finding ways to control the human experience and behaviour in a virtual conversation is the main aim of this thesis.

Alessi and Huang (2000) suggest that virtual humans should be social, emotionally expressive and interactive. That is, virtual humans should give an appropriate response to human's emotional states in terms of speech, facial, and body expression and should take cultural, educational and cognition aspects of an individual into consideration. In order to realize this, a virtual human simulation should integrate a diverse set of artificial intelligent technologies, including speech recognition, natural language understanding and generation, dialog management, non-verbal communication including animated facial expression and body posture, and automated reasoning (Gratch et al., 2002; Swartout, 2006). Extensive research has already been devoted to the development of conversational virtual humans, e.g., in a chatting environment (Ahn et al., 2012), as persuasive agents using body languages (Andre et al., 2011), as intelligent tutors for the domain of negotiation and cultural awareness (Core et al., 2006), as autonomous sensitive listeners (Kokkinara et al., 2011), in turn taking strategies (Ter Maat et al., 2011), and in complex social scenarios involving multiple participants and bystanders (Wang et al., 2013). However, at this moment in time it seems still beyond the state of art to build virtual humans that match the vast diversity and flexibility humans display in natural language communication.

On the other hand, even without matching the full capabilities of human dialog partners, various studies have demonstrated that people do react to their virtual counterpart in a manner they would normally do to other humans (De Melo et al., 2012; Garau et al., 2001; Pertaub et al., 2002; Reeves and Nass, 1996), thereby illustrating the general social effectiveness of virtual humans. For example, Pertaub et al. (2002) found that people with a fear of public speaking reported also anxiety when speaking to a virtual audience. Likewise, Garau et al. (2001) showed that in remote meetings where people were represented by avatars communicated better when the avatars exhibited realistic, task-appropriate eye-gaze behaviour. Also De Melo et al. (2012) found that people disliked negotiating with angry virtual humans and tended to treat them as dominant and uncooperative. Often in these cases, conversations with the virtual human were set within a specific context or followed a defined storyline, making them situational dependent. The advantage of a situational dependent conversation is that it strongly limits the set of anticipated human responses. This in turn makes it easier to build a virtual human that functions appropriately. For example, for a course on mathematics the dialog can be expected to centre on mathematics and learning, and is not expected to include communication related to e.g., travelling to a foreign country. Even applications that do not focus on information exchange, but on emotion modification, such as virtual reality exposure therapy (VRET) for the treatment of social anxiety, position a conversation in a social setting, with possible examples as giving a presentation in front of an audience (Pertaub et al., 2002; Slater et al., 1999), buying an item in a shop (Brinkman et al., 2011), having a job interview (Villani et al., 2012), or going on a blind date (Brinkman et al., 2012). Because of the obvious advantages of using situational dependent communication, the research presented in this thesis used VRET for social anxiety as a case domain.

Social anxiety disorder, also referred to as social phobia, is one of the most common anxiety disorders, estimated to affect 12.1% of the US population (Ruscio et al., 2008), 9.3% of the Dutch population (De Graaf et al., 2012), and 6.7% of the European population (Fehm et al., 2005) during their lifetime. These patients are very sensitive to scrutiny by others and feel embarrassed when they are exposed to social or performance situations such as speaking in public, entering a bar, shopping, having a blind date and undergoing a job interview (American Psychiatric Association, 2013). The disorder is often treated with cognitive behaviour therapy (Fava et al., 2001). The behavioural part of this therapy includes exposure to social situations whereby patients are gradually confronted with more anxiety evoking stimuli. Although exposure in real-life (vivo) is effective (Heimberg et al., 1990, 1998), it also has a number of limitations, such as the limited control of stimuli by the therapist, difficulties in arranging appropriate situations, and the limited willingness of patients to expose themselves to these situations (Garcia-Palacios et al., 2007). Exposing patients in virtual reality, often referred to as VRET, has therefore been put forward as an alternative. Similar to exposure in vivo, exposure in virtual reality confronts patients to anxiety provoking social stimuli in a gradual order, from the least anxiety-evoking situation to the most extreme one. Key difference, of course, is that patients in these virtual environments interact with virtual humans instead of with real humans. Meta-analyses indicate that VRET is as effective as exposure in vivo (Gregg and Tarrier, 2007; Parsons and Rizzo, 2008; Powers and Emmelkamp, 2008) in treating some phobias such as fear of flying and fear of height. Several studies (Anderson et al., 2013, 2005; Harris et al., 2002; Klinger et al., 2005; Robillard et al., 2010) also found a positive effect for exposure in virtual reality for the treatment of social anxiety disorder.

One of the noteworthy benefits of using VRET is that it enables therapists to manipulate and control the feared situation and environment, not only between sessions but also within one single session Emmelkamp (2013). But, controlling anxiety in the case of social phobia requires control on the communication between the human patient and the virtual human(s) in the environment. As this is far from trivial, most studies in this area avoid extensive automated human-virtual human conversations. Instead, they follow situations primarily involving monologues such as in public speaking (Anderson et al., 2005; Harris et al., 2002; Klinger et al., 2005; North et al., 1998, 2002; Pertaub et al., 2001, 2002; Slater et al., 1999), or they use precise scenarios for the communication such as when ordering food in a restaurant or a bar (Brinkman et al., 2008; James et al., 2003), when having a one-way question-answer job interview (Kwon et al., 2009), when shopping in a certain store, or when having a blind date (Brinkman et al., 2012; Ter Heijden and Brinkman, 2011). As mentioned before, the work presented in this thesis builds on this tradition of precise scenarios using situational dependent conversations.

Some research already studied specific anxiety arousing elements for social phobic patients, including body posture (Anderson et al., 2003; Herbelin, 2005; Klinger et al., 2004; Slater et al., 2006) and eye gazing of the conversational partner (Herbelin et al., 2002; Riquier et al., 2002), the kind of narrative text preceding the exposure (Brinkman et al., 2012), and general remarks made by the virtual human (Brinkman et al., 2012). But also more environmental aspects of the virtual world may affect the perceived anxiety in patients. In general, three main elements may be identified that fully characterize a given virtual setting. Taking figure 1 as a representative example, we may distinguish: (1) the surrounding environment, such as tables, picture frames on the wall and televisions; (2) the virtual conversation partner, i.e., the virtual human who talks and listens to the human user (so, the girl in the middle of the picture in figure 1); and (3) virtual bystanders, i.e., the virtual humans that not directly take part in the conversation with the human user, but instead are present in the background of the virtual world, talking for example to each other or interacting with virtual objects. Potentially, all three main elements may provide ways to control the virtual conversation, and therefore are studied in this thesis.

1.1 Research question and hypotheses

Missing insights into how to use the three elements in a virtual world to control the human experience and behaviour in virtual conversation within the setting of VRET for the treatment of social anxiety lead to the main research question of this thesis:

Can and in what way do the virtual surrounding, the behaviour of a virtual dialog partner, and the behaviour of the virtual bystanders have an effect on an individual who is engaged in a conversation with a virtual dialog partner?

In order to answer this main research question, the three elements were empirically studied in four separated studies, each examining their own hypothesis. The first position argued for in this thesis relates to the surrounding environment and how it can affect the virtual conversation. Specifically, the concept of priming is examined for its ability to limit the scope of possible human responses in order to create appropriate replies by a virtual human. Priming can be seen as the incidental activation of a person's knowledge structure which can lead the person to exhibit specific behaviour and attitudes (Bargh, 2006; Bargh

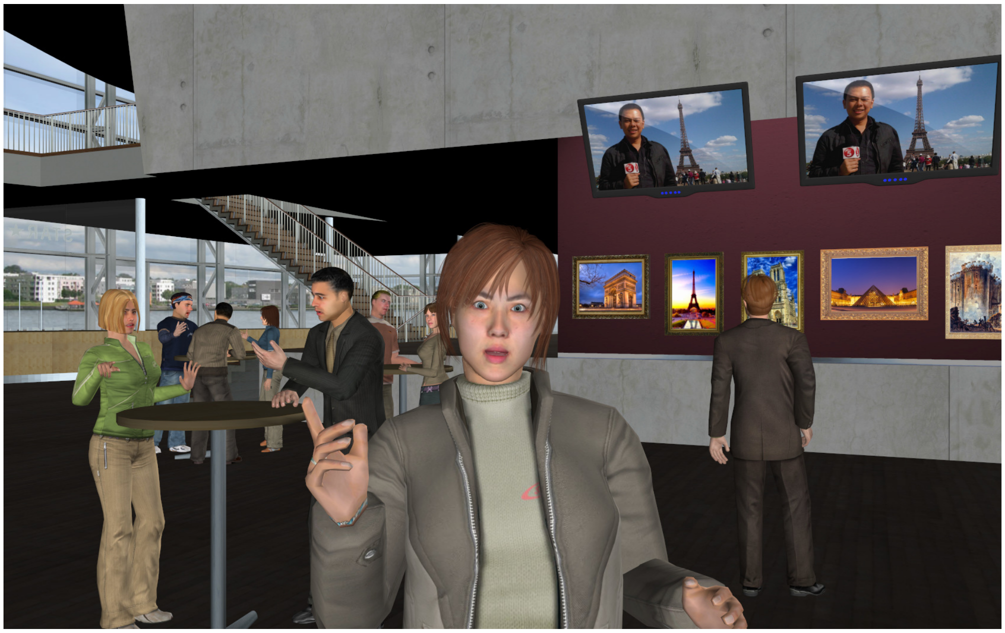


Figure 1.1: Social setting for a virtual conversation including the three key elements studied in this thesis: the surrounding environment, the virtual conversation partner, and virtual bystanders.

et al., 1996). Various studies have examined the concept of priming such as in daily television advertisement (Harris et al., 2009), with colour (Mayr et al., 2009), or with temperature (Williams and Bargh, 2008), and these studies have indicated that indeed priming may influence people's behaviour. These results were used as inspiration to use priming to the benefits of a virtual conversation, i.e., by driving the responses given by a human conversation partner in a specific direction. Ideally, subliminal hints would stimulate people to mention specific keywords, that then can easily be recognised by a computer, and lead to an appropriate reply from a virtual human. In the context of VRET for the treatment of social phobia, the conversational goal is emotion modification, e.g., evoke social anxiety, and not the exchange of information. Therefore, using priming to influence what an individual would say in a conversation has no negative impact on this goal. Thus, the first position argued in this thesis is that priming cues such as videos and pictures can restrict the variety of human responses to match a set of pre-defined keywords, each linked to an appropriate reply from the virtual conversation partner making the flow of a conversation more natural.

The second position put forward in this thesis relates to the effect the virtual human can have on the conversation. More specifically, we argue that the effect of the human perception of the emotion expressed by a virtual human, i.e., the synthetic emotion, may depend on the phase of the conversation. The emphasis of emotion expression largely depends on the application. Some virtual reality applications, such as health coaches (Konstantinidis et al., 2009), need an emotional expression of the virtual human during the speaking phase. Other applications mainly benefits from emotional expressions during the listening phase, such as for a virtual audience in a public speaking environment (Ling et al., 2013; Pertaub et al., 2002). Finally, in some applications emotional expressions are important in both the speaking and listening phases, such as for a conversational partner in job interviews (Brinkman et al., 2012). Studies have investigated how humans perceive virtual human's emotions during the listening (Pertaub et al., 2002; Slater et al., 1999; Wong and McGee, 2012) and speaking phase (MacDorman et al., 2010; Qiu and Benbasat, 2005) separately. To our knowledge, no study has directly compared how synthetic emotions during both speaking and listening in virtual reality are perceived, which would, of course, be relevant when considering a virtual conversation. During the speaking phase, a virtual human talks and simultaneously expresses emotions with both verbal and non-verbal behaviour including facial expression, gaze and head movement, while the listening phase is mainly dominated by non-verbal behaviour to express emotion. This unbalance in channels to express emotion posits that humans may perceive the emotion of a virtual human as more intense in the speaking phase than in the listening phase. As a first step, the work presented in this thesis only looks at the valence dimension of emotion, i.e., positive or negative affect, and therefore the second position defended in this thesis addresses the perception of the valence intensity of an emotion expressed by a virtual human while speaking or listening.

The aim of a virtual conversation as part of VRET for the treatment of social phobia is to elicit anxiety; therefore, next to have synthetic emotions that are correctly perceived, the virtual human should also be able to elicit anxiety. Hence the third position of this thesis relates to humans' responses to and their satisfaction with the virtual conversation. Affective feedback plays an important role in a conversation and it may cause supportive or defensive responses from a listener in human-human communication (Gibb, 1961). Similar results were found in virtual worlds (Burleson and Picard, 2007; De Melo et al., 2012; Pertaub et al., 2002). For example, Burleson and Picard (2007) found that a system with a virtual character that provided affective support reduced frustration of less confident users. Maldonado et al. (2005) found that a positive emotion expressed by a co-learner enhanced student's learning gains and enjoyment. Pertaub et al. (2002) found that a negative audience elicited a significantly higher level of anxiety in their group of participants as compared to a neutral or positive audience. This thesis therefore argues that a virtual human can elicit positive or negative affect in a human conversation partner, and as such, may affect satisfaction towards the conversation. This ability of virtual humans may allow therapists to have more options to control the anxiety stimuli.

The fourth and last position of this thesis relates to the third element in a virtual social setting, which are the virtual bystanders. These characters, although present, do not directly take part in the conversation. They can be regarded as intentional or unintentional observers of the social interaction. The effect bystanders may have has extensively been studied in the past. For example, Asch (1951) demonstrated their effect on people's judgement, whereby people have the tendency to comply with the majority view of bystanders. Another effect bystanders can have is known as the social facilitation tendency (Geen, 1989) in that people perform better in the presence of others on a well-trained task and worse on an untrained task. Finally, there is the phenomenon known as the bystander effect (Darley and Latane, 1968). This refers to the observation that the likelihood a person would help a victim is inversely related to the number of presented bystanders. In addition, observing others in a social context is also an important way for people to learn as is postulated by the social cognitive theory (Bandura, 1997, 2001) and is a central idea when it comes to the development of people's self-efficacy (Bandura, 1997), i.e., people's belief in their own ability to perform a certain task. The above mentioned effects of bystanders have also been studied in virtual reality. Kozlov and Johansen (2010) and Slater et al. (2013) were able to demonstrate that the bystander effect can be replicated in a virtual environment, whereas Park and Catrambone (2007) demonstrated the ability to replicate the social facilitation phenomenon in virtual reality. Furthermore, observing virtual humans perform certain actions, e.g., physical exercises, has also been suggested to affect the observers' self-efficacy about these actions (Fox and Bailenson, 2009). Thus, the fourth position defended in this thesis is that virtual bystanders can affect a person's beliefs and behaviour during a virtual conversation.

To conclude this section, from the main research question and the four main tenets introduced, it is now possible to derive the following hypotheses that are tested in this thesis:

1. Priming pictures and videos increase the chance that individuals use specific keywords in their answers when having a human-virtual human conversation.

2. The virtual human's expressed valence is perceived as more intense in the speaking phase than in the listening phase.

3. By expressing a positive or negative emotion, a virtual human can elicit a

corresponding emotional state in a human conversation partner and affect the satisfaction towards the conversation.

4. Virtual bystanders can affect a person's beliefs and behaviour during a virtual conversation.

1.2 Methodology and thesis structure

In order to test the first hypothesis, regarding priming people to mention a specific keyword in their answer, two experiments were conducted. The first experiment examined whether priming worked in a real life conversation. Once that was established, the second experiment was conducted to demonstrate that this effect could be replicated in virtual reality. In the first experiment, twenty participants 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 experimenter who asked the questions. The second experiment shared the same experimental setting, but was carried out in virtual reality instead of in the real world. Twenty participants were asked to answer questions from a virtual human when they were exposed to priming material, i.e., videos and images in the virtual environment, before and/or during the conversation session. In both experiments the participants' answers were analysed in terms of the number of times they mentioned a word from the target set. The empirical studies and their results are described in Chapter 2.

Chapter 3 describes an empirical experiment testing the second hypothesis regarding the perception of a conversational virtual human. As part of this experiment, two validation studies of the stimuli were first conducted: validating the emotion expressed in voice and whether the intensity differences in the nonverbal emotional behaviour during listening and speaking could be distinguished. For the main experiment, 24 participants (12 Chinese, 12 non-Chinese) were recruited and asked to rate the valence of seven different emotional expressions (ranging from negative over neutral to positive during the speaking and listening phase) of a Chinese virtual lady who also spoke only in Chinese. The perceived valence in the speaking and listening phase was analysed, as well as the effect of cultural difference on perceived valence.

In order to test the third hypothesis, a within-subjects empirical study with six conditions using the same Chinese virtual lady as in Chapter 3 was conducted. For each condition, the virtual lady's emotions in the listening and speaking phase were different, including positive, neutral and negative emotions. Twenty-four Chinese participants were recruited and exposed to all the six conditions, with a different conversation topic in each condition. A presence questionnaire, the dialog satisfaction questionnaire and the Self-Assessment Manikin questionnaire were administered after each conversation with the virtual human. During the conversation, participants' dialog length and physiological data such as heart rate and skin conductance were recorded. The experiment and its results are described in Chapter 4.

The last hypothesis regarding the effect of virtual bystanders was tested with twenty-six participants exposed to four virtual English lessons to practise speaking in English. The virtual students in the classroom represented the virtual bystanders in this social setting. Each lesson consisted of two phases; in the first phase, the virtual English teacher asked four virtual peer students questions about everyday life issues, while in the second phase, the participants were requested to answer four questions from the virtual English teacher. The four lessons were created by manipulating two within-subjects variable: (1) the bystanders' attitude towards the virtual peer speakers, and (2) the bystanders' attitude towards the participants when they were answering questions of the teacher. The virtual students' attitude, which could either be positive or negative, was expressed mainly by facial expressions and by comments whispered between the bystanders. A questionnaire measured the participants' anxiety, self-efficacy and beliefs after each session. To measure physical arousal, physiological data such as heart rate and skin conductance were again collected during the exposure. To measure avoidance behaviour, the length of the participants' answers was recorded and analysed. This study and its results are presented in Chapter 5.

The conclusions that can be drawn from the studies presented in this thesis are discussed in Chapter 6, including also the main contributions of this research and suggestions for future research.

Bibliography

Ahn, J., Gobron, S., Garcia, D., Silvestre, Q., Thalmann, D., and Boulic, R. (2012). An NVC Emotional Model for Conversational Virtual Humans in a 3D Chatting Environment. *Lecture Notes in Computer Science*, 7378:47–57.

Alessi, N. E. and Huang, M. P. (2000). Evolution of the Virtual Human: From term to potential application in psychiatry. *Cyberpsychology & Behavior*, 3(3):321–326.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders*. Author, Washington, DC, 5th edition.

Anderson, P. L., Price, M., Edwards, S. M., Obasaju, M. a., Schmertz, S. K., Zimand, E., and Calamaras, M. R. (2013). Virtual reality exposure therapy for social anxiety disorder: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, 81(5):751–760.

Anderson, P. L., Rothbaum, B. O., and Hodges, L. F. (2003). Virtual Reality Exposure in the Treatment of Social Anxiety. *Cognitive And Behavioral Practice*, 10(3):240–247.

Anderson, P. L., Zimand, E., Hodges, L. F., and Rothbaum, B. O. (2005). Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and Anxiety*, 22(3):156–158.

Andre, E., Bevacqua, E., Heylen, D., Niewiadomski, R., Pelachaud, C., Peters, C., Poggi, I., and Rehm, M. (2011). Non-verbal Persuasion and Communication in an Affective Agent. In *Emotion-Oriented Systems Cognitive Technologies*, pages 585–608. Springer.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. *Groups, Leadership, and Men. S*, pages 222–236.

Balcisoy, S., Torre, R., Ponder, M., Fua, P., and Thalmann, D. (2000). Augmented reality for real and virtual humans. *Computer Graphics International 2000, Proceedings*, pages 303–307.

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. Worth Publishers.

Bandura, A. (2001). Social cognitive theory of mass communication. *Media Psychology*, 3(3):265–299.

Bargh, J. A. (2006). 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., and Burrows, L. (1996). Automaticity of social behavior: direct effects of trait construct and stereotype-activation on action. *Journal of Personality and Social Psychology*, 71(2):230–244.

Brinkman, W.-P., Hartanto, D., Kang, N., De Vliegher, D., Kampmann, I. L., Morina, N., Emmelkamp, P. M. G., and 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*, pages 1099–1102.

Brinkman, W.-P., Hattangadi, N., Meziane, Z., and Pul, P. (2011). Design and Evaluation of a Virtual Environment for the Treatment of Anger. In Richir, S. and Akihiko, S., editors, *Proceedings of Virtual Reality International Conference (VRIC 2011)*, pages 6–8, Laval, France.

Brinkman, W.-P., Van der Mast, C. A. P. G., and 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*, pages 83–95.

Broekens, J., Harbers, M., Brinkman, W.-P., Jonker, C., Van den Bosch, K., and Meyer, J.-J. (2012). Virtual reality negotiation training increases negotiation knowledge and skill. In *IVA'12 Proceedings of the 12th international conference on Intelligent Virtual Agents*, pages 218–230.

Burleson, W. and Picard, R. W. (2007). Gender-specific approaches to developing emotionally intelligent learning companions. *Intelligent Systems*, 22(4):62–69.

Core, M., Traum, D., Lane, H. C., Swartout, W. R., Marsella, S., Gratch, J., and Van Lent, M. (2006). Teaching negotiation skills through practice and reflection with virtual humans. *Simulation*, 82:685–701.

Darley, J. M. and Latane, B. (1968). Bystander Intervention in Emergencies - Diffusion of Responsibility. *Journal of Personality and Social Psychology*, 8(4p1):377–383.

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

De Melo, C., Carnevale, P., and Gratch, J. (2012). The Effect of Virtual Agents' Emotion Displays and Appraisals on People's Decision Making in Negotiation. *Intelligent Virtual Agents*, pages 53–66.

Dow, S., Mehta, M., Harmon, E., MacIntyre, B., and Mateas, M. (2007). Presence and engagement in an interactive drama. *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, pages 409–416.

Emmelkamp, P. M. G. (2013). Behavior Therapy with Adults. In Lambert, M. J., editor, *Bergin and Garfield's Handbook of Psychotherapy and Behavior*, pages 343–392. John Wiley & Sons.

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

Fehm, L., Pelissolo, A., Furmark, T., and Wittchen, H.-U. (2005). Size and burden of social phobia in Europe. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology*, 15(4):453–462.

Foutz, S., Ancelet, J., Hershorin, K., and Danter, L. (2012). Responsive Virtual Human Museum Guides: Summative Evaluation. Technical report, Institute for Learning Innovation.

Fox, J. and Bailenson, J. N. (2009). Virtual Self-Modeling: The Effects of Vicarious Reinforcement and Identification on Exercise Behaviors. *Media Psychology*, 12(1):1–25.

Garau, M., Slater, M., Bee, S., and Sasse, M. A. (2001). The Impact of Eye Gaze on Communication Using Humanoid Avatars.

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

Geen, R. G. (1989). Alternative conceptions of social facilitation. In Paulus, P., editor, *Psychology of Group Influence Hillsdale*, pages 15–51. Lawrence Erlbaum Associates, Mahwah, NJ.

Gibb, J. R. (1961). Defensive Communication. *Journal of Communication*, 11(3):141–148.

Gratch, J., Rickel, J., Andre, E., Cassell, J., Petajan, E., and Badler, N. I. (2002). Creating interactive virtual humans: Some assembly required. *Intelligent Systems, IEEE*, 17(4):54–63.

Gregg, L. and Tarrier, N. (2007). Virtual reality in mental health: a review of the literature. *Social Psychiatry and Psychiatric Epidemiology*, 42(5):343–354.

Harris, J. L., Bargh, J. A., and Brownell, K. D. (2009). Priming effects of television food advertising on eating behavior. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*, 28(4):404–413.

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

Heimberg, R. G., Dodge, C. S., Hope, D. A., Kennedy, C. R., Zollo, L. J., and Becker, R. E. (1990). Cognitive Behavioral Group Treatment for Social Phobia - Comparison with a Credible Placebo Control. *Cognitive Therapy and Research*, 14(1):1–23.

Heimberg, R. G., Liebowitz, M. R., Hope, D. A., Schneier, F. R., Holt, C. S., Welkowitz, L. A., Juster, H. R., Campeas, R., Bruch, M. A., Cloitre, M., Fallon, B., and Klein, D. F. (1998). Cognitive behavioral group therapy vs phenelzine therapy for social phobia - 12-week outcome. *Archives of General Psychiatry*, 55(12):1133–1141.

Herbelin, B. (2005). *Virtual reality exposure therapy for social phobia.* PhD thesis, Louis Pasteur University.

Herbelin, B., Riquier, F., Vexo, F., and 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*, pages 1–10.

Hill, R., Gratch, J., Marsella, S., Rickel, J., Swartout, W. R., and Traum, D. (2003). Virtual humans in the mission rehearsal exercise system. *Kunstliche Intelligenz*, 4(3):5–10.

James, L. K., Lin, C.-Y., Steed, A., Swapp, D., and Slater, M. (2003). Social anxiety in virtual environments: results of a pilot study. *Cyberpsychology & Behavior*, 6(3):237–243.

Kenny, P. G., Parsons, T. D., Gratch, J., and Rizzo, A. A. (2008). Evaluation of Justina: A Virtual Patient with PTSD.

Klinger, E., Bouchard, S., Legeron, P., Roy, S., Lauer, F., Chemin, I., and 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., and Nugues, P. (2004). Virtual Reality Exposure in the Treatment of Social Phobia. *Studies in Health Technology and Informatics*, 99:91–119.

Kokkinara, E., Oyekoya, O., and Steed, A. (2011). Modelling selective visual attention for autonomous virtual characters. *Computer Animation and Virtual Worlds*, 22(4):361–369.

Konstantinidis, E. I., Hitoglou-Antoniadou, M., Luneski, A., Bamidis, P. D., and Nikolaidou, M. M. (2009). Using affective avatars and rich multimedia content for education of children with autism. *Proceedings of the 2nd International Conference on PErvsive Technologies Related to Assistive Environments - PETRA '09*, pages 1–6.

Kopp, S., Gesellensetter, L., Kramer, N. C., and Wachsmuth, I. (2005). A conversational agent as museum guide Design and Evaluation of a Real-World Application. . In Panayiotopoulos, T., Gratch, J., Aylett, R. S., Ballin, D., Olivier, P., and Rist, T., editors, *Intelligent Virtual Agents 2005*, pages 329–343, Kos, Greece.

Kozlov, M. D. and Johansen, M. K. (2010). Real Behavior in Virtual Environments: Psychology Experiments in a Simple Virtual-Reality Paradigm Using

Video Games. *Cyberpsychology Behavior and Social Networking*, 13(6):711–714.

Kwon, J., Alan, C., and Czanner, S. (2009). A study of visual perception: social anxiety and virtual realism. In *Proceeding SCCG '09 Proceedings of the 25th Spring Conference on Computer Graphics*, pages 167–172.

Ling, Y., Nefs, H. T., Qu, C., Heynderickx, I., and Brinkman, W.-P. (2013). The effect of perspective on presence and space perception. *PLoS ONE*, 8(11):e78513.

MacDorman, K. F., Coram, J. A., Ho, C.-C., and Patel, H. (2010). Gender differences in the impact of presentational factors in human character animation on decisions in ethical dilemmas. *Presence: Teleoperators and Virtual Environments*, 19(3):213–229.

Maldonado, H., Lee, J.-e. R., Brave, S., Nass, C., Nakajima, H., Yamada, R., Iwamura, K., and Morishima, Y. (2005). We Learn Better Together : Enhancing eLearning with Emotional Characters. In *Computer Supported Collaborative Learning 2005: The Next 10 Years!*, pages 408–417. Lawrence Erlbaum Associates, Mahwah, NJ.

Mateas, M. and Stern, A. (2003). Facade: An experiment in building a fully-realized interactive drama. *Game Developers Conference. Game Design Track.*

Mayr, S., Hauke, R., Buchner, A., and Niedeggen, M. (2009). No evidence for a cue mismatch in negative priming. *Quarterly journal of experimental psychology (2006)*, 62(4):645–652.

Muller, T. J., Heuvelink, A., van den Bosch, K., and Swartjes, I. (2012). Glengarry Glen Ross: Using BDI for Sales Game Dialogues. *Proceedings, The Eighth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment.*

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

Park, S. and Catrambone, R. (2007). Social facilitation effects of virtual humans. *Human Factors*, 49(6):1054–1060.

Parsons, T. D. and 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.

Pertaub, D.-P., Slater, M., and Barker, C. (2001). An experiment on fear of public speaking in virtual reality. *Studies in Health Technology and Informatics*, 81:372–378.

Pertaub, D.-P., Slater, M., and 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. and Emmelkamp, P. M. G. (2008). Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders*, 22(3):561–569.

Qiu, L. and Benbasat, I. (2005). Online consumer trust and live help interfaces: The effects of text-to-speech voice and three-dimensional avatars. *International Journal of Human-Computer Interaction*, 19(1):37–41.

Reeves, B. and Nass, C. (1996). *The Media Equation*. Cambridge University Press.

Riquier, F., Stankovic, M., and Chevalley, A. F. (2002). Virtual gazes for social exposure: Margot and Snow White. In *Proceedings of the 1st. International Workshop on Virtual Reality Rehabilitation*.

Robillard, G., Bouchard, S., Dumoulin, S., Guitard, T., and 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.

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

Slater, M., Pertaub, D.-P., Barker, C., and Clark, D. M. (2006). An experimental study on fear of public speaking using a virtual environment. *Cyberpsychology & Behavior*, 9(5):627–633.

Slater, M., Pertaub, D.-P., and Steed, A. (1999). Public speaking in virtual reality: facing an audience of avatars. *IEEE Computer Graphics and Applications*, 19(2):6–9.

Slater, M., Rovira, A., Southern, R., Swapp, D., Zhang, J. J., Campbell, C., and Levine, M. (2013). Bystander responses to a violent incident in an immersive virtual environment. *PLoS ONE*, 8(1):e52766.

Swartout, W. R. (2006). Virtual humans. In *Proceedings of the National Conference on Artificial Intelligence*, volume 2, pages 1543–1545, Boston, MA; United States.

Ter Heijden, N. and 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 Maat, M., Truong, K. P., and Heylen, D. (2011). How Agents' Turn-Taking Strategies Influence Impressions and Response Behaviors. *Presence: Teleoperators and Virtual Environments*, 20(5):412–430.

Traum, D., Rickel, J., Gratch, J., and Marsella, S. (2003). Negotiation over Tasks in Hybrid Human Agent Teams for Simulation Based Training.

Vandeventer, J. and Barbour, B. (2010). Sammi: A 3-Dimensional Virtual Human Information Kiosk. In *ACM SE '10 Proceedings of the 48th Annual Southeast Regional Conference*, Oxford, MS, USA.

Villani, D., Repetto, C., Cipresso, P., and Riva, G. (2012). May I experience more presence in doing the same thing in virtual reality than in reality? An answer from a simulated job interview. *Interacting with Computers*, 24(4):265–272.

Wang, Z., Lee, J., and Marsella, S. (2013). Multi-party, multi-role comprehensive listening behavior. *Autonomous Agents and Multi-Agent Systems*, 27(2):218–234.

Williams, L. E. and Bargh, J. A. (2008). Experiencing physical warmth promotes interpersonal warmth. *Science (New York, N.Y.)*, 322(5901):606–607.

Wong, J. W.-E. and McGee, K. (2012). Frown More, Talk More: Effects of Facial Expressions in Establishing Conversational Rapport with Virtual Agents. In *IVA'12 Proceedings of the 12th international conference on Intelligent Virtual Agents*, pages 419–425.

Chapter 2

The Virtual Surroundings

The effect of priming:

manipulating pictures and videos on a dialog scenario in a virtual environment

This chapter has been published as:
Chao Qu, Willem-Paul Brinkman, Pascal Wiggers, Ingrid Heynderickx (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.

Having a free speech conversation with virtual humans in a virtual environment can be desirable in virtual reality applications such as virtual reality exposure 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 reality 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 a virtual human 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.

2.1 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 et al., 2001, 2004; Krijn et al., 2004b; 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 VRET (Krijn et al., 2004a).

In VRET for individuals with social phobia, interaction between a patient and a virtual human, i.e., a virtual human needs to arouse a certain level of social anxiety (Robillard et al., 2010). Regulating the response of the virtual human 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 et al., 2010). To alleviate this workload, integration of a keyword based dialog manager into a VRET system has been proposed (Ter Heijden et al., 2010). Compared to the human control, maintaining the conversation between patient and virtual human 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 and Brinkman, 2011).

In order to use a dialog manager, usually the dialog content needs to be predefined. Figure 2.1 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.

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.

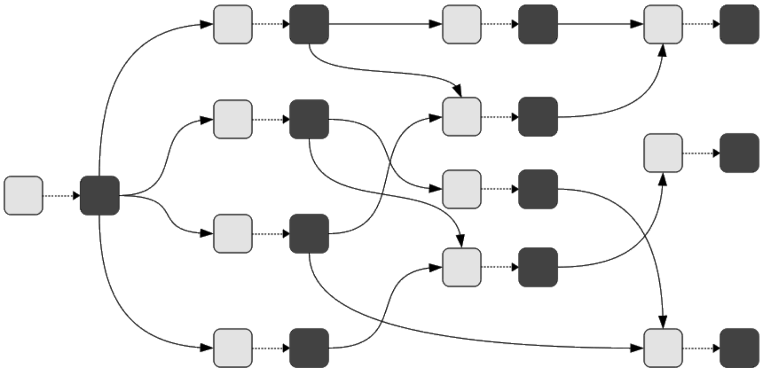


Figure 2.1: Example of a dialog structure; the dark blocks are the computer's utterances, while the light blocks represent the user's responses. All possible responses are pre-defined.

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 et al., 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 a virtual human, 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 et al., 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 a virtual human, 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 a VR environment to influence user’s responses in a conversation with a virtual human?

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 a virtual human.

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.

2.2 Theoretical Background

2.2.1 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 et al., 2012) and in general 6.7% of the European population (Fehm et al., 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 are afraid of embarrassing themselves in social situations, they feel uncomfortable and try to avoid being exposed to social situations (American Psychiatric Association, 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).

2.2.2 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 et al., 2001; Walshe et al., 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 and Chambless, 1995; Taylor, 1996; Gould et al., 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 et al., 2003; Rothbaum et al., 1996), fear of height (Krijn et al., 2004b; Rothbaum et al., 1995), fear of special insects (Carlin et al., 1997; Garcia-Palacios et al., 2002; Botella et al., 2005), and treatment of post-traumatic stress disorder (Difede and Hoffman, 2002). Recent meta-analyses indicate that VRET is as effective as exposure in vivo (Gregg and Tarrier, 2007; Parsons and Rizzo, 2008; Powers and 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 virtual humans 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 virtual humans (North et al., 1998; Slater et al., 1999; Pertaub et al., 2001; Harris et al., 2002; North et al., 2002; Pertaub et al., 2002; Anderson et al., 2005; Klinger et al., 2005; Slater et al., 2006b) or ordering food in a restaurant or a bar (James et al., 2003; Klinger et al., 2004). The variety in virtual human's behavior is then usually limited to the body posture (Anderson et al., 2003; Herbelin, 2005; Klinger et al., 2004; Slater et al., 2006a) and eye gazing (Riquier et al., 2002; Herbelin et al., 2002). Moreover, verbal responses of the virtual human 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 and Brinkman, 2011; Brinkman et al., 2012).

2.2.3 Speech Processing and Dialog Manager

Using speech recognition to analyze what the patient is saying and automatically selecting an appropriate virtual human 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 and 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 and Martin, 2000), not to mention free speech processing.

The conversational agents such as real estate agents (Cassell et al., 1999), e-retail (McBreen and Jack, 2001) and automated phone reservation systems (McTear et al., 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 and Brinkman, 2011). The virtual humans 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 virtual human.

2.2.4 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 et al., 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 and 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 et al., 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 et al., 2006; Sperber et al., 1979; Rosch, 1975; Williams and Bargh, 2008; Harris et al., 2009; Yap et al., 2011), such as daily television advertisement priming (Harris et al., 2009), masked picture priming with precise time control (Marzouki et al., 2007, 2008), colour priming (Mayr et al., 2009) and temperature priming (Williams and Bargh, 2008). Among these studies, some priming experiments are related to virtual reality (Pena et al., 2009; Nunez and 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.

2.2.5 Presence

The concept of presence contains several very different facets. Generally it covers two sub-concepts: physical (or spatial) presence and social presence (IJsselsteijn et al., 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 et al., 2005). Social presence refers to the feeling of being together with another person (Biocca et al., 2001) or the illusion of sharing the same physical space (Riva et al., 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 et al., 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 et al., 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 and 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 and 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 et al., 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-virtual human 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 et al., 2001). This questionnaire is widely used (Alsina-Jurnet and Gutierrez-Maldonado, 2010; Freire et al., 2010; Krijn et al., 2004a; Ling et al., 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.

2.2.6 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.

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

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

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.

2.3 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.

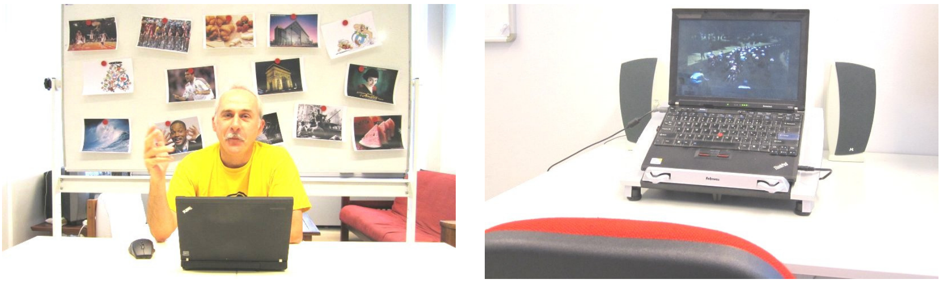
2.3.1 Experiment Design

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 2.1.

Table 2.1: 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 and 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 2.2(b). During the conversation, pictures were attached on a whiteboard right behind the interviewer, as shown in figure 2.2(a). 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



(a) with picture

(b) with video

Figure 2.2: Priming with Picture/Video in Real World.

without informing the participant who was watching the videos in the other room.

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.

2.3.2 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.

2.3.3 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 2.2(b)). 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.

2.3.4 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.

2.3.5 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 ($F(1,19) = 7.12, p = .015$), 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 2.3 and table 2.2 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., 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.

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.

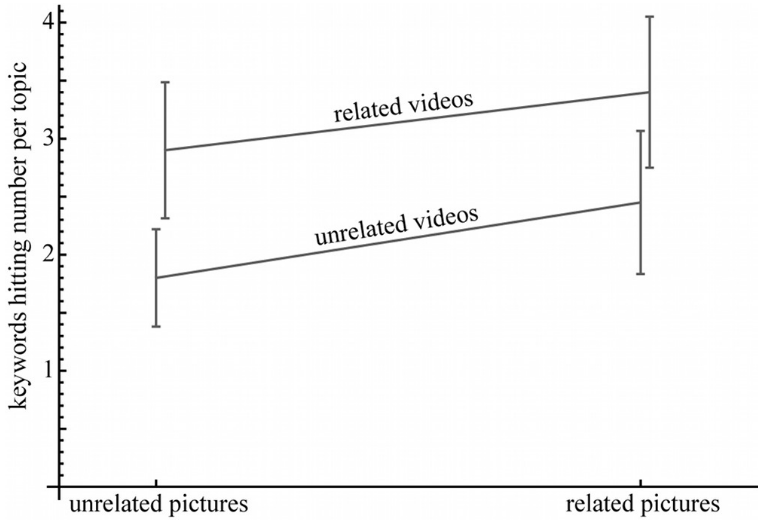


Figure 2.3: Effect of priming with pictures and videos in the real world based on the mean value of the number of keywords hit per topic and per priming condition, including the 95% confidence interval.

Table 2.2: Means, Standard Deviations and Bounds in terms of KHN of different conditions in study 1

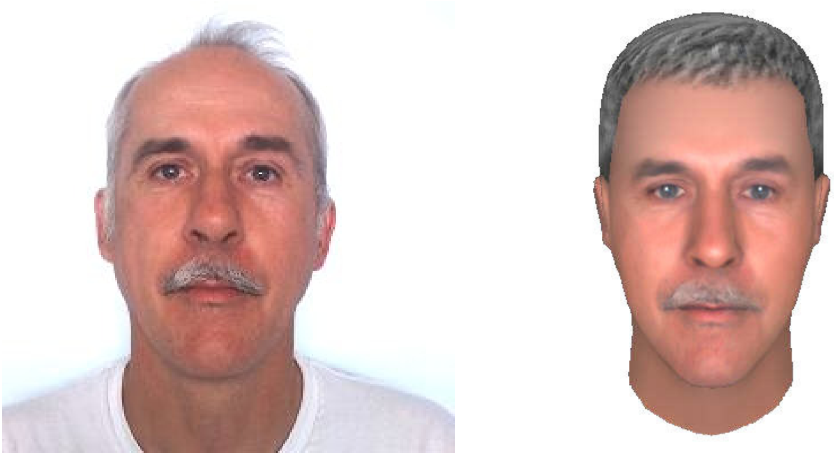
			95% Confidence Interval	
Condition	Mean	Std. Deviation	Lower Bound	Upper Bound
unrelated picture & unrelated video	1.80	0.89	1.38	2.22
unrelated picture & related video	2.90	1.25	2.31	3.49
related picture & unrelated video	2.45	1.32	1.83	3.07
related picture & related video	3.40	1.39	2.75	4.05

2.4 Study 2, Human-virtual human 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 a virtual human 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 a virtual human, as shown in figure 2.4.

2.4.1 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



(a) Photo (b) Virtual Human

Figure 2.4: Experimenter and his virtual human.

embedded in a virtual picture frame, and the videos were embedded in a virtual television (figure 2.6(b)). The experiment had again a two-by-two within-subject design, with four counterbalanced conditions in a reduced Latin square (Denes and Keedwell, 1974) and four randomly assigned conversation topics, similar to study 1.

2.4.2 Materials

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

2.4.3 Procedure

A small room created with 3Ds MAX was chosen for this experiment, as shown in figure 2.5. During the experiment, participants sat on a big sofa in the middle of the room *(A)* while answering the questions of the virtual human who was sitting in front of them on another sofa *(B)*. Right behind the virtual human was a white wall *(C)* where the pictures were hanging. The participant was able to see the pictures when he or she faced the virtual human. The television

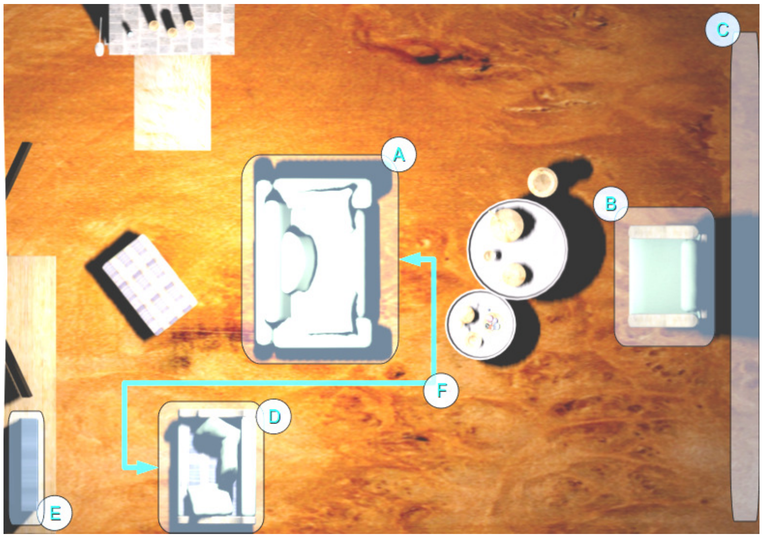


Figure 2.5: Top View of the Virtual Room for the Experiment.

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 virtual human, the dialog started.

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

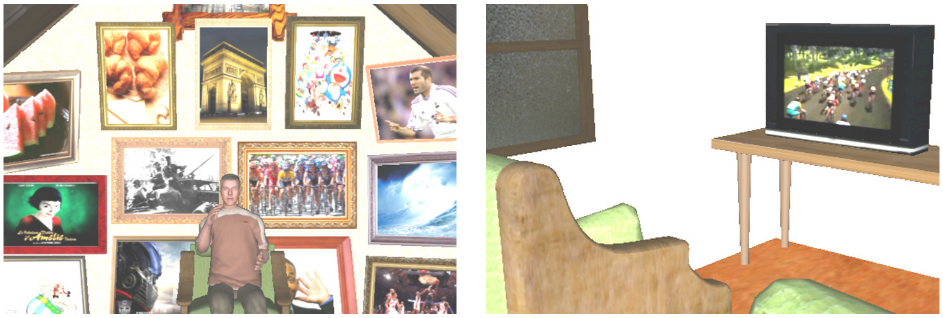
Participants wore an eMagin[5] Z800 Head-Mounted Display (HMD) to observe the virtual world. The eMagin Z800 is a USB-powered immersive display device, with a resolution of 800 × 600 pixels. With a build-in 360-degree advanced head tracker, the participant could turn his or her head freely to perceive all the pictures and videos (figure 2.6(a), 2.6(b)). The whole conversation with the virtual human 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.,

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

[3] http://www.facegen.com

[4] http://www.worldviz.com

[5] http://www.emagin.com



(a) with picture

(b) with video

Figure 2.6: Priming with Picture/Video in Virtual World.

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.

2.4.4 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.

2.4.5 Results

Data Analysis of Study 2

Table 2.3 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

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

applied in other studies (Ter Heijden and Brinkman, 2011; Ling et al., 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.

Table 2.3: Means and standard deviations for Igroup Presence Questionnaire

			95% Confidence Interval	
Subscales	Mean	Std. Deviation	Lower Bound	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 ($F(1, 19) = 0.33, p = .577$) 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 2.7 and table 2.4 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.

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

`ipq/data.php`

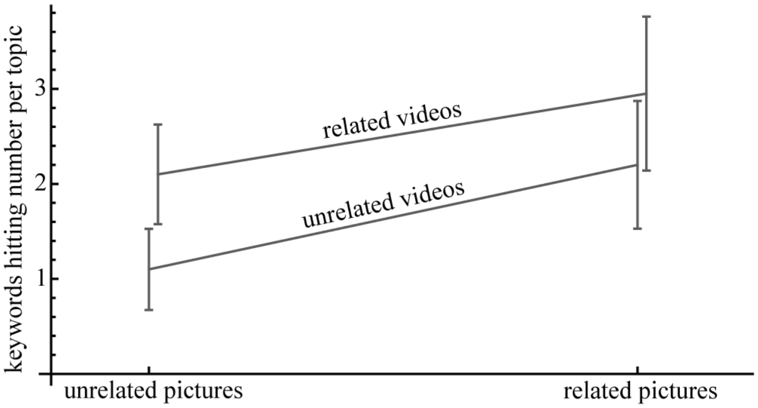


Figure 2.7: Effect of priming with pictures and videos in the virtual world based on the mean value of the number of keywords hit per topic and per priming condition, including the 95% confidence interval.

Table 2.4: 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

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.

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.

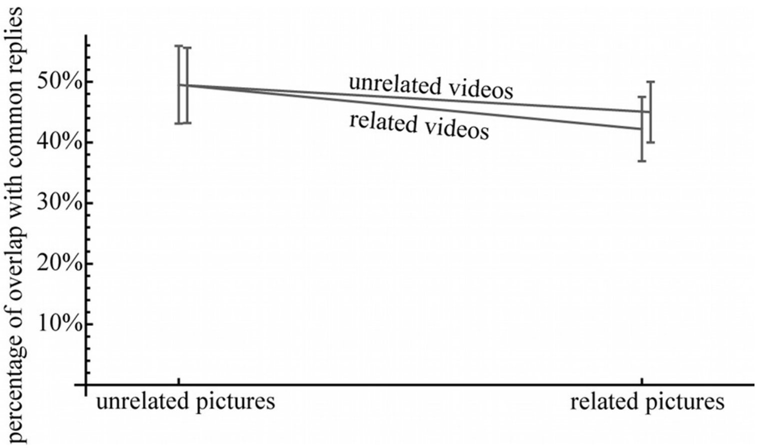


Figure 2.8: Mean percentage of overlap between the responses given with the top 40% of common responses for each of the priming conditions. The whiskers represent the 95% confidence interval.

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 2.8 and table 2.5 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 2.5: Means, Standard Deviations and Bounds of overlap between the responses given with the top 40% of common responses in different conditions

			95% Confidence Interval	
Condition	Mean	Std. Deviation	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

2.5 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 a virtual human 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 a virtual human.

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 virtual human 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 et al., 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 to the same VR world multiple times. Even if the patient talks about the same topic with the virtual human, 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 dialog with a virtual human, 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 virtual human. 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 virtual humans, 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 virtual human looks at them, or whether the virtual human shows a negative attitude towards them (Clark and McManus, 2002). These negative social interactions are likely to motivate patients not to look directly at the virtual human (Chen et al., 2002; Horley et al., 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 virtual human. 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.

Bibliography

Alsina-Jurnet, I. and 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.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR.* American Psychiatric Pub., Washington, DC.

Amir, N. (2003). 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., and 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., and 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., and Hodges, L. F. (2003). Virtual Reality Exposure in the Treatment of Social Anxiety. *Cognitive And Behavioral Practice*, 10(3):240–247.

Anderson, P. L., Zimand, E., Hodges, L. F., and Rothbaum, B. O. (2005). Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and Anxiety*, 22(3):156–158.

Bargh, J. A. (2006). 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., and Burrows, L. (1996). Automaticity of social behavior: direct effects of trait construct and stereotype-activation on action. *Journal of Personality and Social Psychology*, 71(2):230–244.

Biocca, F. A., Harms, C., and 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., and Rey, B. (2005). Mixing realities? An application of augmented reality for the treatment of cockroach phobia. *Cyberpsychology & Behavior*, 8(2):162–171.

Brinkman, W.-P., Hartanto, D., Kang, N., De Vliegher, D., Kampmann, I. L., Morina, N., Emmelkamp, P. M. G., and 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*, pages 1099–1102.

Brinkman, W.-P., Van der Mast, C. A. P. G., and 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*, pages 83–95.

Brinkman, W.-P., Van der Mast, C. A. P. G., Sandino, G., Gunawan, L. T., and Emmelkamp, P. M. G. (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.

Busscher, B., De Vliegher, D., Ling, Y., and 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., and 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. W., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson, H. H., and 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*, number May in CHI '99, pages 520–527. ACM Press.

Chen, Y. P., Ehlers, A., Clark, D. M., and 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. and McManus, F. (2002). 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., and 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. and Keedwell, A. D. (1974). *Latin squares and their applications.* Academic Press, New York.

Difede, J. and Hoffman, H. G. (2002). Virtual reality exposure therapy for World Trade Center Post-traumatic Stress Disorder: a case report. *Cyberpsychology & Behavior*, 5(6):529–535.

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

Emmelkamp, P. M. G., Bruynzeel, M., Drost, L., and Van der Mast, C. A. P. G. (2001). Virtual reality treatment in acrophobia: a comparison with exposure in vivo. *Cyberpsychology & Behavior*, 4(3):335–339.

Emmelkamp, P. M. G., Krijn, M., Hulsbosch, a. M., De Vries, S., Schuemie, M. J., and Van der Mast, C. A. P. G. (2002). Virtual reality treatment versus exposure in vivo: a comparative evaluation in acrophobia. *Behaviour Research and Therapy*, 40(5):509–516.

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

Fehm, L., Pelissolo, A., Furmark, T., and Wittchen, H.-U. (2005). Size and burden of social phobia in Europe. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology*, 15(4):453–462.

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

Feske, U. and Chambless, D. L. (1995). 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., and 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., and 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., and Botella, C. M. (2002). Virtual reality in the treatment of spider phobia: a controlled study. *Behaviour Research and Therapy*, 40(9):983–993.

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

Gregg, L. and Tarrier, N. (2007). Virtual reality in mental health: a review of the literature. *Social Psychiatry and Psychiatric Epidemiology*, 42(5):343–354.

Harris, J. L., Bargh, J. A., and Brownell, K. D. (2009). Priming effects of television food advertising on eating behavior. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*, 28(4):404–413.

Harris, S. R., Kemmerling, R. L., and North, M. M. (2002). 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., and Neerincx, M. A. (2012). Automatic mechanisms for measuring subjective unit of discomfort. *Annual Review of Cybertherapy and Telemedicine*, 181:192–196.

Herbelin, B. (2005). *Virtual reality exposure therapy for social phobia.* PhD thesis, Louis Pasteur University.

Herbelin, B., Riquier, F., Vexo, F., and 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*, pages 1–10.

Horley, K., Williams, L. M., Gonsalvez, C., and Gordon, E. (2003). 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. and Alder, M. D. (1998). Introducing MegaHAL. *In nemlap3, Conll98 workshop on human-computer conversation, ACL*, pages 271–274.

IJsselsteijn, W. A., De Ridder, H., Freeman, J., and Avons, S. E. (2000). Presence: concept, determinants, and measurement. In *SPIE: Human Vision and Electronic Imaging V*, volume 3959, pages 520–529.

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

James, L. K., Lin, C.-Y., Steed, A., Swapp, D., and Slater, M. (2003). Social anxiety in virtual environments: results of a pilot study. *Cyberpsychology & Behavior*, 6(3):237–243.

Jurafsky, D. and Martin, J. H. (2000). *Speech and Language Processing.* Prentice Hall Series in Artificial Intelligence. Prentice Hall.

Klinger, E., Bouchard, S., Legeron, P., Roy, S., Lauer, F., Chemin, I., and 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., and Nugues, P. (2004). 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., and De Ligny, C. D. (2004a). 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., and Biemond, R. (2004b). Virtual reality exposure therapy of anxiety disorders: a review. *Clinical Psychology Review*, 24(3):259–281.

Larsson, S. (2000). 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., and 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., and Heynderickx, I. (2013). The relationship between individual characteristics and experienced presence. *Computers in Human Behavior*, 29(4):1519–1530.

Lombard, M., Ditton, T. B., Crane, D., Davis, B., Gil-Egui, G., Horvath, K., Rossman, J., and 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., and Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, 101(1):3–33.

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

Marzouki, Y., Grainger, J., and Theeuwes, J. (2008). Inhibition of return in subliminal letter priming. *Acta psychologica*, 129(1):112–120.

Mayr, S., Hauke, R., Buchner, A., and Niedeggen, M. (2009). No evidence for a cue mismatch in negative priming. *Quarterly journal of experimental psychology (2006)*, 62(4):645–652.

McBreen, H. M. and 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., and Liu, X. (2005). Handling errors and determining confirmation strategies - An object-based approach. *Speech Communication*, 45(3):249–269.

Muhlberger, A., Wiedemann, G. C., and 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., and 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., and 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. and 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*, pages 101–108, New York, USA. ACM Press.

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

Parsons, T. D. and 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.

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

Pertaub, D.-P., Slater, M., and Barker, C. (2001). An experiment on fear of public speaking in virtual reality. *Studies in Health Technology and Informatics*, 81:372–378.

Pertaub, D.-P., Slater, M., and 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. and Emmelkamp, P. M. G. (2008). Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders*, 22(3):561–569.

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

Redi, J., Liu, H., Zunino, R., and Heynderickx, I. (2011). Interactions of visual attention and quality perception. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7865, page 26.

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

Riquier, F., Stankovic, M., and 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., and 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., and 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., and North, M. M. (1995). 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. A., Kessler, G. D., and 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., and Kessler, R. C. (2008). 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. and Slater, M. (2005). From presence to consciousness through virtual reality. *Nature reviews. Neuroscience*, 6(4):332–339.

Schubert, T., Friedmann, F., and Regenbrecht, H. (2001). The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and 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., Pistrang, N., and Sanchez-Vives, M. V. (2006a). A virtual reprise of the Stanley Milgram obedience experiments. *PLoS ONE*, 1(1):1–10.

Slater, M., Pertaub, D.-P., Barker, C., and Clark, D. M. (2006b). An experimental study on fear of public speaking using a virtual environment. *Cyberpsychology & Behavior*, 9(5):627–633.

Slater, M., Pertaub, D.-P., and Steed, A. (1999). Public speaking in virtual reality: facing an audience of avatars. *IEEE Computer Graphics and Applications*, 19(2):6–9.

Sperber, R. D., McCauley, C., Ragain, R. D., and Weil, C. M. (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. and 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., and 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*, pages 978–994.

Von Der Putten, A. M., Klatt, J., Ten Broeke, S., McCall, R., Kramer, N. C., Wetzel, R., Blum, L., and 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., and Kim, S. I. (2005). Virtually driving: are the driving environments "real enough" for exposure therapy with accident victims? An explorative study. *Cyberpsychology & Behavior*, 8(6):532–537.

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. and Bargh, J. A. (2008). Experiencing physical warmth promotes interpersonal warmth. *Science (New York, N.Y.)*, 322(5901):606–607.

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

Yap, D.-F., So, W.-C., Melvin Yap, J.-M., Tan, Y.-Q., and Teoh, R.-L. S. (2011). Iconic gestures prime words. *Cognitive Science*, 35(1):171–183.

Chapter 3

The Virtual Dialog Partner I

The effect of emotion and culture:

an empirical study on human perception of an emotional virtual dialog partner

This chapter has been published as:
Chao Qu, Willem-Paul Brinkman, Yun Ling, Pascal Wiggers, Ingrid Heynderickx (2013). Human perception of a conversational virtual human: an empirical study on the effect of emotion and culture. *Virtual Reality*, 17(4), 307-321.

Virtual reality applications with virtual humans, such as virtual reality exposure therapy, health coaches, and negotiation simulators, are developed for different contexts and usually for users from different countries. The emphasis on a virtual human's emotional expression depends on the application; some virtual reality applications need an emotional expression of the virtual human during the speaking phase, some during the listening phase and some during both speaking and listening phases. Although studies have investigated how humans perceive a virtual human's emotion during each phase separately, few studies carried out a parallel comparison between the two phases. This study aims to fill this gap, and on top of that, includes an investigation of the cultural interpretation of the virtual human's emotion, especially with respect to the emotion's valence. The experiment was conducted with both Chinese and non-Chinese participants. These participants were asked to rate the valence of seven different emotional expressions (ranging from negative to neutral to positive during speaking and listening) of a Chinese virtual lady. The results showed that there was a high correlation in valence rating between both groups of participants, which indicated that the valence of the emotional expressions was as easily recognized by people from a different cultural background as the virtual human. In addition, participants tended to perceive the virtual human's expressed valence as more intense in the speaking phase than in the listening phase. The additional vocal emotional expression in the speaking phase is put forward as a likely cause for this phenomenon.

3.1 Introduction

To create a feeling of being “present” in virtual reality is essential to the success of many virtual reality applications such as training (Broekens et al., 2011), coaching (Rizzo et al., 2011), therapy (Brinkman et al., 2012), and games (Isbister, 2006). A feeling of being “present” in virtual reality may be achieved by making the virtual reality environment as natural as possible. Human-computer interaction, including human-virtual human interaction, is inherently natural and social (Reeves and Nass, 1996), and so, is an essential component in the realism of the virtual environment. Without proper behavior of the virtual human, users may not be able to “suspend disbelief” and the effectiveness of the virtual reality application will decrease.

Considering the importance of emotion in human-human communication, emotion may also help people to establish a better relationship with virtual human (Reeves and Nass, 1996). As Picard et al. (2001) argues, without some emotional skills, machines will not appear intelligent when interacting with people. Therefore, multiple technologies to give virtual human the abilities of generating human acceptable expressions have been developed in recent decades (Ersotelos and Dong, 2008).

Different applications require different levels of emphasis on how the virtual human express their emotions. Even when implemented in only part of the application, emotional expressions can be effective. In a health coach application, for example, the virtual human might mainly need to speak to motivate the user, and emotional expressions during speaking are most important. In a virtual reality exposure therapy for fear of public speaking, the virtual human only needs to listen, and so, emotional expressions while listening are most important. In some applications, such as for a role playing game or a negotiation simulator, the full range of speaking and listening is used and might benefit from emotional expressions. Studies on generating and evaluating the emotional agents normally only focus on either listening (Slater et al., 1999; Wong and McGee, 2012) or speaking (MacDorman et al., 2010; Qiu and Benbasat, 2005). Studies that do include both speaking and listening, (e.g., Core et al., 2006; Broekens et al., 2012a; Link et al., 2006) focus mainly on the conversation and communication as a whole, and do not separately investigate the speaking and listening phase of the whole conversation. To our knowledge, no study has directly compared the impact of emotional expressions during speaking and listening in virtual reality. In the current study, the virtual human’s valence state was manipulated while she was speaking and listening from negative to positive, and the impact on the participants’ perception was examined.

Besides the difference between listening and speaking, culture might also be an important factor for a designer to consider as many applications are used all across the world nowadays. Especially for some virtual reality applications, such as virtual reality exposure therapy for patients with social phobia (Brinkman et al., 2012), it is crucial to understand how people with a different cultural background perceive the affective behavior of virtual humans. Several studies have already focused on the effect of cultural differences on evaluating virtual human's emotions. For example, Jack et al. (2012) showed that facial expressions of emotion are culture specific. However, Yun et al. (2009) found that cultural background has little effect on emotion perception. Kleinsmith et al. (2006) evaluated cultural impact on perception of emotion and found that emotions are both universal and culturally specific. Therefore, similar to the studies for perceiving emotional expressions of real humans, the universality of emotion perception of virtual human seems also still inconclusive. In addition, most research is only limited to the investigation of head-only virtual human with facial expressions and far less research is devoted to emotional expressions from a 3D virtual human which expresses its emotional state also via gaze, head movement or voice intonation.

In summary, this study involves two research questions: (1) whether emotional expressions of virtual human are perceived differently depending on the cultural background of the perceiver, and (2) whether a person is more perceptive to emotional expressions in one of the two phases (the speaking phase and listening phase) or whether a person treats these two phases as equally important when rating the virtual human's emotions? To answer these research questions, we designed a virtual human representing a Chinese lady at an age around 25. She had the ability to show multiple emotional states in multiple non-verbal and verbal ways: i.e., through facial expression, head movement, gaze and voice intonation. During the listening phase, the virtual human's emotional behavior was expressed by non-verbal communication only, while during the speaking phase, the emotional behavior was expressed by both verbal (i.e., intonation) and non-verbal communication. To avoid a possible emotional bias from the content of the conversation, a relatively neutral topic, i.e., conference attendance, was selected in this experiment. Petrushin (1999) pointed out that humans are not perfect in decoding manifest emotions such as anger and happiness in voice intonation only. Therefore, as a first step, only three basic emotional valence states (positive, neutral and negative) were used in this study. In order to test the effect of cultural influence on the perception of the emotions expressed by the virtual human, two groups of participants were recruited: from the same culture as the virtual human and from other cultures. We chose to compare Chinese versus non-Chinese participants, because it is known from cultural models (Hofstede, 2001) that the difference in cultural values is significant between these two groups, and since two of the authors experienced these differences while living in Europe. Moreover the background of these authors facilitated the recruitment of Chinese participants.

Based on knowledge available in the literature, we envision the following two hypotheses related to our research questions.

HYPOTHESIS 1: Individuals with the same cultural background as the virtual human perceive the valence state of the virtual human differently from individuals with a different cultural background.

Especially, as the virtual human was speaking Chinese, participants with a different cultural background could not understand what the virtual human said during verbal communication. Hence, participants with a different cultural background from the virtual human are expected to perceive her emotion differently from participants with the same cultural background.

HYPOTHESIS 2: The virtual human's expressed valence is perceived as more intense in the speaking phase than in the listening phase.

Since the speaking phase also allows including verbal expression of the emotion, it seems likely that compared to the listening phase, the emotion expressed in the speaking phase is perceived as more intense.

The rest of the paper is structured as follows. Section 2 provides theoretical background on how a virtual human can express emotion through facial expression, gaze, head movements, and voice intonation. In addition, it discusses cultural differences in emotion recognition and various emotional models, needed to understand the rest of the paper. Section 3 provides a description of the apparatus, validation of the stimuli material and the procedure of the experiment, and its results are presented in section 4. Finally, in section 5 the findings of the study are discussed and conclusions are drawn.

3.2 Theoretical Background

No matter what roles virtual humans play in a virtual world, they need to elicit an anthropomorphic interaction with their human users. This requires vast knowledge of various human aspects including facial expression, gaze, head movement, voice expression and their cultural difference in order to make the virtual human believable, responsive and interpretable.

3.2.1 Facial Expression of a Virtual Human

Facial expression is one of the options to express human emotion, and as such plays a substantial role in depicting human characters. Started in the early 70s 80s (Parke, 1972; Platt and Badler, 1981), face modeling and animation have been a continuous research topic for many years. From early 2000s, more flexible emotion representations were created with MPEG-4-based facial animation (Tsapatsoulis et al., 2002). Recent advances in facial animation that allow to produce a rich set of effects on synthetic humans already had their impact on the industry (Ersotelos and Dong, 2008).

Multiple approaches have been proposed to create naturally looking facial expressions; they can be categorized as follows: (1) simulation or physically based models, which try to model the anatomical structure of the face as well as the underlying dynamics (Kahler et al., 2001; Lee et al., 1995; Waters, 1987), (2) performance driven models, which reassemble frames from video footage or motion capture data of a real person to yield the desired facial expression (Brand, 1999; Bregler, 1997; Chuang and Bregler, 2002; Ezzat et al., 2004; Litwinowicz and Williams, 1994)), and (3) parameterized based models, which assign weights to the vertices of meshes representing the face, such that during animation the vertices are moved according to the weights (Cohen and Massaro, 1993; Parke, 1974; Zhang et al., 2006). Considering the high computational load required for the simulation or physically based models and the high costs for the motion capture equipment needed for performance driven models, we decided to choose an easily repeated facial expression animation based on a parameterized model for this study.

3.2.2 Head Movement and Gaze of a Virtual human

Besides facial expressions, also head and eye movements were implemented in the virtual human used in our experiment. Head movements and eye gaze are two important sources of emotional feedback in interaction (Cassell and Thorisson, 1999; Lee and Marsella, 2012; Ruttkay and Pelachaud, 2005). They are essential to embody interactive conversational systems (Cassell et al., 1994) and it is relatively simple to create primarily nods and glances towards or away from the user. Still the correct timing is essential (Cassell and Thorisson, 1999). Research of Lance et al. (2008) and Lee et al. (2009) show how head movements and gaze can be embedded into a virtual character.

3.2.3 Voice Expression of a Virtual human

Along with the non-verbal emotional expressions, emotion can also be expressed by voice intonation when the virtual human is talking. Speech was once considered as the main channel to carry most, or even all, the necessary information in a conversation (Ochsman and Chapanis, 1974). This idea has been countered by a growing body of research on believable, life-like embodied conversational agents (Bates, 1994). Still, the importance of the voice in emotion expression cannot be denied (Scherer, 1995). Many studies have investigated emotional effects in voice and speech (Bailenson et al., 2006; Petrushin, 1999; Scherer, 2003), and emotion expressed in the voice of virtual humans (Cerezo and Baldassarri, 2008; Moridis and Economides, 2012). The intonation of the voice was therefore also considered as an important aspect of the virtual human's emotional expression in this study.

3.2.4 Cultural Difference

Culture, like age, gender, posture and context, is one of the many factors affecting emotion expression (Picard, 1998). A long-time question in the study of human emotion is the extent to which emotional expressions are universal or culturally determined (Elfenbein et al., 2007). Cultural background may influence the rate of emotion recognition (Matsumoto, 2002). When an expresser of an emotion and the perceiver of the emotion have the same cultural background, the perceiver's recognition rate is found to be higher than when the expresser and perceiver have a different cultural background (Elfenbein, 2003; Elfenbein and Ambady, 2002; Elfenbein et al., 2007). However, Darwin (1872) and Tomkins (1962, 1963) argue that universal emotions do exist, studies also show universality in the facial expression of emotion and its perception, and attribute only little effect of cultural background on emotion perception from facial expressions (Ekman, 1994; Ekman and Friesen, 1971; Ekman et al., 1987; Matsumoto, 2002, 2007).

The question of impact of cultural background can be extended to human-virtual human interaction. Although various studies show that people can correctly identify emotions expressed by embodied agents in general (Bartneck, 2001; Schiano et al., 2000), how good this performance is retained in different cultures needs to be considered. Clear indications support the statement that culture can shape the expression and interpretation of emotions (Keltner and Ekman, 2000). Culture as a factor has also been studied in the interaction with computers. For example, Dotsch and Wigboldus (2008) and Brinkman et al. (2011) have found a difference in emotional reaction to a virtual human with ethnic appearance that match or did not match the person's ethnicity. Endrass et al. (2011) show that in German and Japanese cultures, the user's perception of an agent conversation can be enhanced by a culturally prototypical performance of gestures and body postures. Kleinsmith et al. (2006) worked on the cross-cultural difference of recognizing affect from virtual human's body posture and suggest to consider culture as one specific factor for the implementation of agents. Jan et al. (2007) mention that in Arabian and US American cultures, gaze, proximity and turn-taking behavior are all culture related. These results reveal that participants perceive behavior that is in line with their own cultural background differently from behavior that is typical for a different cultural background. In the work presented in this paper, cultural background is considered as a variable which is expected to influence how people perceive the emotional expression of the virtual human.

3.2.5 Dimensional Emotion Model

Although for facial expressions six universal basic emotions exist (Ekman et al., 1992), for language people's categorization of verbal labels to describe their everyday life emotions vary between languages and cultures (Russell, 1991). Instead of placing these expressed emotions in categories, i.e., a discrete emotional approach, others suggest placing them in a multi-dimensional space, i.e., a dimensional approach (Fox, 2008). Three broad dimensions have often been proposed to describe affect (Mehrabian and Russell, 1974): i.e., valence, arousal and dominance. Valence is variously referred to as positive and negative affect or as pleasant and unpleasant feelings. The arousal dimension ranges emotions from deep sleep to frenetic excitement. Dominance focuses on the expression of social control and aggression, and varies between submissive and dominant (Schroder, 2004). Compared to the discrete emotional approach, the dimensional approach often uses subjective reports of feelings as its main dependent variable. As such, it has a strong empirical base. Support for the existence of these dimensions has come from research into subjective reports, physiological responses, neural circuits, and cognitive appraisal (Barrett, 2006; Fox, 2008). Furthermore, Wierzbicka (1995) and Church and Katigbak (1998) also investigated the cross-cultural universality of the emotional dimensions. Their results showed the universality of the valence and arousal dimensions. The study presented in this paper focuses on the valence dimension only. Although participants were asked to rate the virtual human's emotion on all the three dimensions, only the valence dimension was used for data analysis.

3.3 Experiment

3.3.1 Participants

Twelve Chinese (7 female and 5 male) and twelve non-Chinese (5 female and 7 male) students from the Delft University of Technology participated in the experiment. Their age ranged from 24 to 38 years with a mean of 27.8 ($SD =$ 3.4) years. All participants were naive with respect to the hypotheses. Written informed consent forms were obtained from all the participants. The experiment was approved by the university ethic committee.

3.3.2 Creating the virtual human

Although Cowell and Stanney (2003) found that users generally prefer to interact with a youthful character matching their ethnicity, they found no significant preference for character gender. Furthermore Kulms et al. (2011) showed that actual behavior and its evaluation are more important for the evaluation than gender stereotypes. Therefore, a Chinese virtual lady aged around 25 years was specially created for this study.

The model of the virtual human was created by $FaceGen$ and $3Ds\ MAX$. All main factors which were considered to contribute to emotion expression were combined; the virtual human's facial expression, her head and eye movements and her voice intonation were manipulated to express emotion during the conversation. To create facial expressions, an easily repeated facial expression animation method was used. This method rigged the face mesh into 22 action units with 18 features (Gratch et al., 2002), where each feature was an anchor point attached to a set of vertices of the face. A model for the face dynamics that was able to control the intensity of the expression, its onset, peak and decay was defined. As such, the virtual human had the ability to show any intensity and any combination of the six basic Ekman facial expressions Ekman and Friesen (1978). The validation of this approach was shown by Broekens et al. (2012b). By setting the values for the three emotional dimensions (i.e., the valence, arousal and dominance), and for the expression duration, any emotion could be expressed by the virtual human. The facial expressions from neutral to negative or from neutral to positive, used by the virtual human in our experiment are shown in figure 3.1.

The participants were asked to judge the emotional state of the virtual human, and so, there was no interaction between the participant and the virtual human. The participant was told that the scene contained a virtual lady talking with a human, but that the human voice was removed. Therefore, problems related to timing (i.e., whether the virtual human should or should not show an expression at a certain point of time) were avoided, and the participant could focus on the emotional behavior of the virtual human herself.

Seven conditions were included in the experiment, all varying in the emotional states of the virtual human. Since the scenario was conversation based, two continuously alternating phases could be identified, i.e., one in which the virtual human was speaking and one in which she was listening. These phases allowed the virtual human to express her emotion differently in the two phases. In the speaking phase, the virtual human used voice and non-verbal communication to express her emotions, while in the listening phase the virtual human only used non-verbal communication to express her emotions. Three emotional states were created for both phases (i.e., positive, neutral and negative state),

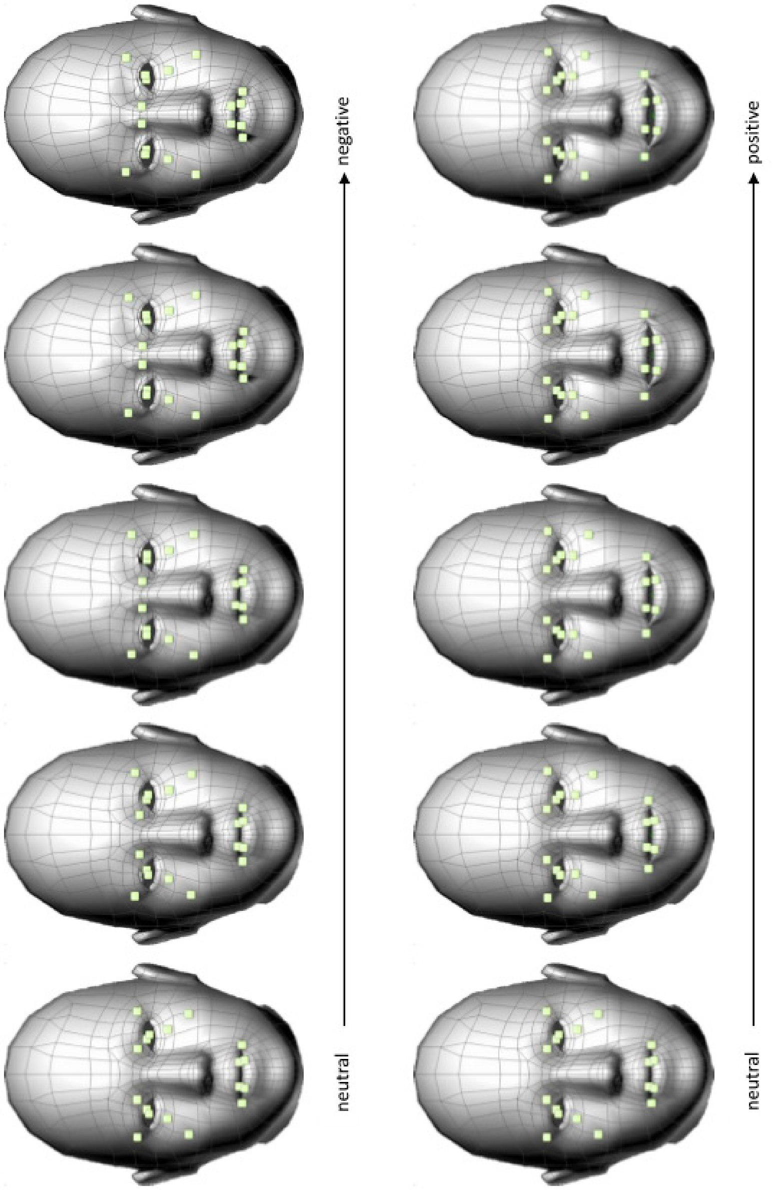


Figure 3.1: Emotions expressed by moving some action units (i.e., the small squares in the figure) of the face mesh. Left column: emotions changing from neutral to negative; right column: emotions changing from neutral to positive.

and they formed the basis for the seven different conditions, shown in figure 3.2. As the combination of positive (negative) listening and negative (positive) speaking included contradictory emotional information of the virtual human in the speaking and listening phase, these combinations were considered unnatural, and so, were excluded from the experiment. Taking the neutral attitude in both the speaking and listening phase as the baseline, it was expected that participants would give a higher valence score when the virtual human responded positively either in the listening or speaking phase. Assuming that there would be no interaction between the speaking and listening phase and that both phases would have a similar impact on the expressed valence intensity, the seven conditions could be ordered into five groups: highly negative (`S-L-`), lowly negative (`S-L0`, `S0L-`), neutral (`S0L0`), lowly positive (`S+L0`, `S0L+`) and highly positive (`S+L+`). If the intensity of the expressions with a negative or positive valence would be equal, these five groups could be projected on a single valence scale as is done in figure 3.2 (shown as the predicted valence value axis). Comparing the actual valence values obtained in the experiment to the predicted valence values would make it possible to study hypothesis 2 about the experience of the valence intensity in the two phases of the conversation.

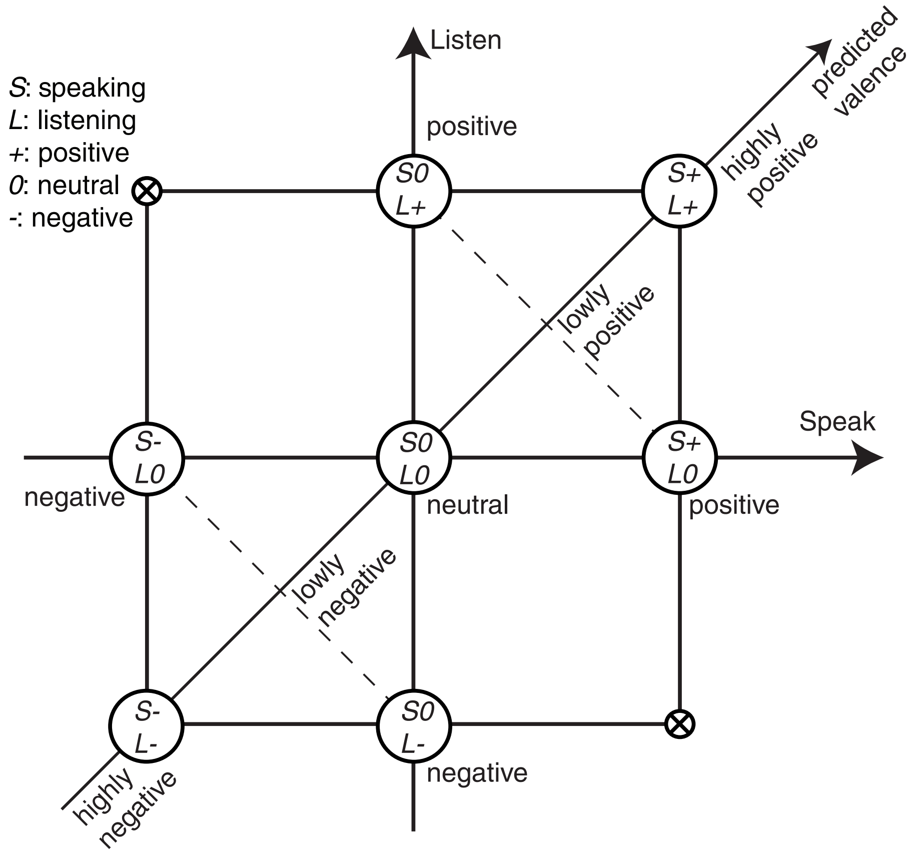
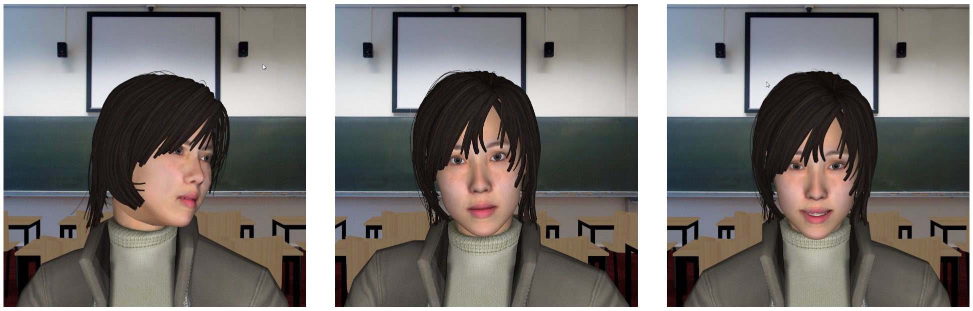


Figure 3.2: Seven conditions, existing of combinations of an emotional state in the speaking and listening phase of a conversation, as used in the experiment and their corresponding predicted valence intensity.

The participants were asked to sit in front of the virtual human (displayed only above her chest on a computer screen), right at the place where the virtual human's "conversational partner" would sit. With this set up, the participants could well perceive the virtual human's emotional state, expressed by her vocal expression, facial expression, eyes and head movements. When expressing a positive emotional state, the virtual human would show a happy facial expression, and once in a while would nod her head to agree with her conversation partner. Her eyes would mainly look at her conversational partner, only occasionally look away (figure 3.3(c)). When expressing a negative emotional state the virtual human would have an angry facial expression and would continuously look away showing limited interest in her conversation partner (figure 3.3(a)). The intensity of both the positive (happy) and negative (angry) emotional expression was evaluated in a previous study (Broekens et al., 2012b) to ensure that they both could be identified by individuals. The neutral expression was the default facial expression of FaceGen, with the six Ekman basic emotion (Ekman et al., 1992) parameters set to zero and with all other morph modifiers removed when generating the face model.



(a) Negative: angry facial expression, only looking at her conversation partner at the beginning, gradually losing interest and starting to look around.

(b) Neutral: neutral facial expression while constantly looking at her conversation partner with slight eye movements.

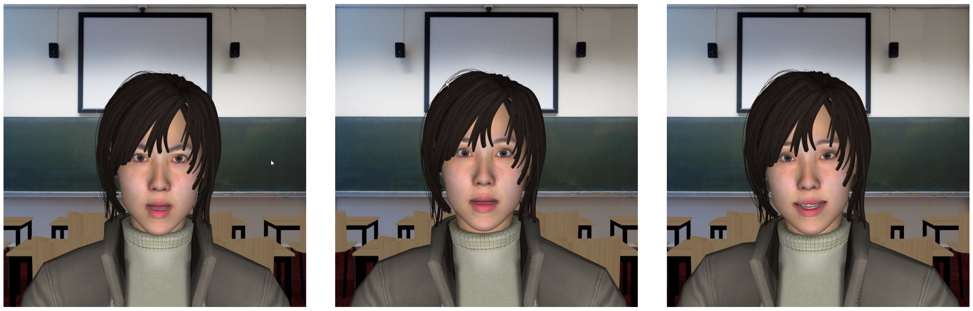
(c) Positive: happy facial expression while constantly looking at her conversation partner, showing some slight eye movements, and occasionally nodding her head.

Figure 3.3: Different emotional states of the virtual human in her listening phase

In the speaking phase, the virtual human would look directly at her conversation partner. In the negative speaking condition she would have a negative facial expression (figure 3.4(a)), while in the positive speaking condition she would have a positive facial expression (figure 3.4(c)). In addition, speech with either a negative or positive intonation was added to the virtual human.

3.3.3 Emotion Validation

As mentioned already above, for the speaking phase, verbal communication was added to the virtual human. The voice of the virtual human was recorded in



(a) Negative: angry facial expression while looking at her conversation partner, and speaking with a negative voice intonation.

(b) Neutral: neutral facial expression while constantly looking at her conversation partner, and speaking with a neutral voice intonation.

(c) Positive: happy facial expression while constantly looking at her conversation partner, and speaking with a positive voice intonation

Figure 3.4: Different emotional states of the virtual human in her speaking phase

Chinese by a Chinese linguistics student. Her voice was recorded 3 times, each time expressing a different emotional state: positive, neutral and negative. A small separate study, in which 6 Chinese participants, 3 male and 3 female with an average age of 27 ($SD = 0.5$) years, were asked to rate the valence of the recorded voice on a scale from 1 (negative) to 9 (positive), showed that the emotion in the recorded voice was indeed perceived as intended, $F(2, 10) = 25.29, p < .001$. The negative voice was significantly lower than the neutral voice, $t(5) = 3.87, p = .012$, and the positive voice, $t(5) = 6.52, p < .001$. Further, the positive voice was significantly higher than the neutral voice, $t(5) = 3.61, p = .015$. The means and standard deviations of the scores on the positive, the neutral and the negative voice were $M = 7.8, SD = 1.9$; $M = 5.7, SD = 2.0$; $M = 1.7, SD = 0.8$, respectively.

Making a fair comparison between the listening and speaking phase requires that the intensity of the non-verbal communication is similar in both phases. For example, the virtual human's facial and body expression in the lowly negative speaking phase and lowly negatively listening phase (see figure 3.2) should have a similar impact on the valence intensity. To test this, an additional small study was conducted. In this study twelve participants, 5 male and 7 female with an average age of 27 years ($SD = 1.8$) were presented simultaneously with two video clips of the virtual human including both the listening and speaking phase. Half of the participants were Chinese. The participants were asked to rate how easily they could see the difference between the two videos on a scale from very easy (0) to very difficult (100). The participants were explicitly asked not to rate the valence, but only the easiness with which differences were perceived, representing the intensity of the emotion. The videos were presented without sound. The participants were asked to rate 12 pairs in total (S-L0/S0L0, S0L-/S0L0, S+L0/S0L0, S0L+/S0L0, S-L-/S+L+, S-L-/S0L0, S+L+/S0L0, S0L0/S0L0, S+L0/S+L0, S-L0/S-L0, S0L+/S0L+, S0L-/S0L-), presented to each participant in a different random order. Before they rated the pairs, the participants were shown all the possible behaviors of the virtual human so that they could establish an overall frame of reference.

The first step of the analysis was to see whether the more intense stimuli were easier to distinguish from the neutral reference video (S0L0) and whether the positive and negative videos were equally distinctive. Therefore, a MANOVA with repeated measures was conducted with the intensity of the video stimuli (high versus low intensity) and the valence direction (positive versus negative) as independent variables. The analysis was conducted on the rating for highly positive (S+L+/S0L0) and negative (S-L-/S0L0) videos, and the mean rating for lowly positive (S+L0/S0L0 and S0L+/S0L0) and negative (S-L0/S0L0 and S0L-/S0L0) videos across the speaking and listening phase. The analysis found a significant main effect ($F(1, 11) = 21.91, p = .001$) for intensity, in that the highly positive or negative videos ($M = 32, SD = 17$) were rated as easier to be distinguished than the lowly positive or negative videos ($M = 44, SD = 15$). Also a significant ($F(1, 11) = 15.63, p = .002$) main effect was found for direction. The positive videos ($M = 25, SD = 15$) were rated as more easily to be distinguished from the neutral video than the negative videos ($M = 50, SD = 23$). The analysis found no significant ($F(1, 11) = 1.60, p = .23$) two-way interaction effect, which suggests that the two main effects were constant across the conditions.

The next analysis focused on the question whether, compared to the neutral reference video, the positive or negative differences in the listening or speaking phase were equally distinguishable, and whether this was the same for the positive and negative videos. Therefore, a second MANOVA with repeated measures was conducted with the valence direction and the phase (speaking versus listening) as independent variables. The analysis used the rating for lowly positive speaking (S+L0/S0L0) and lowly positive listening (S0L+/S0L0) phase, and the rating for the lowly negative (S-L0/S0L0) speaking and lowly listening (S0L-/S0L0) phase. The analysis again revealed that the positive videos ($M = 28, SD = 16$) were significantly ($F(1, 11) = 16.91, p = .002$) rated as more easily to be distinguished than the negative videos ($M = 59, SD = 24$) from the neutral reference video. No significant difference was found between the listening and speaking phase ($F(1, 11) = 0.14, p = .71$), and also no significant two-way interaction effect was found ($F(1, 11) = 0.44, p = .52$). Figure 3.5 shows the videos with their predicted valence and the estimated valence. The latter is the z-score of the rating for the video subtracted from the rating of the neutral reference video (S0L0/S0L0) whereby the rating of negative videos was multiplied by -1. Both the two lowly negative and the two lowly positive video are positioned closely together. In other words the intensity of the non-verbal communication seems similar in the listening and speaking phase. Furthermore, because of the significant difference in rating between negative and positive videos, the neutral reference video seems to be positioned closer to the negative videos than to the positive videos. As illustrated in figure 3.5 the predicted and estimated valence values for the videos do not follow a linear function, but rather a cubic function. By using a fitted inverted cubic function, the intensity weighted predicted valence values for the videos were calculated from the estimated valence values, thereby creating values of intended valence intensity to be compared with the perceived valence rating of videos later in the paper.

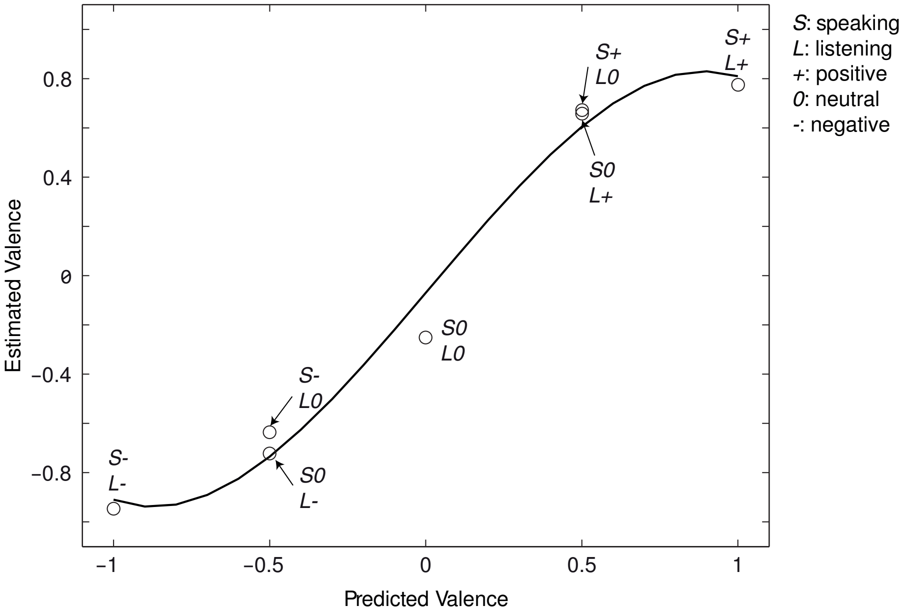


Figure 3.5: Predicted valence plotted against the estimated valence fitted with a cubic function.

3.3.4 Measurements

There are various ways to quantitatively measure the three emotional dimensions (i.e., valence, arousal and dominance). To ensure the reliability of the emotion measurement, two subjective self-reporting instruments were included in this study: the Self-Assessment Manikin Questionnaire (SAM) (Lang, 1995) and the AffectButton (AFB) (Broekens and Brinkman, 2009, 2013).

The SAM questionnaire consists of a series of manikin figures to judge the affective quality (figure 3.6). As a nonverbal rating system, the SAM questionnaire represents the intensity value of the three dimensions of emotion:

valence, arousal and dominance (Lang, 1995). The first row of SAM manikin figures ranges from unhappy to happy on the valence dimension. The second row represents the arousal dimension, ranging from relaxed to excited. The last row ranges from dominated to controlling, representing the dominance dimension. When instructed on how to use the SAM questionnaire according to the detailed explanation, provided in the instruction manual of Lang et al. (2008), participants can select one of the nine figures on each row to express their feelings about the emotional stimulus. The manikin figures were taken from the PXLab (Irtel, 2007). Various studies show that the SAM questionnaire accurately measures emotional reactions to imagery (Lang et al., 1999; Morris, 1995), sounds (Bradley and Lang, 2007), robot gesture expression (Haring et al., 2011), etc.

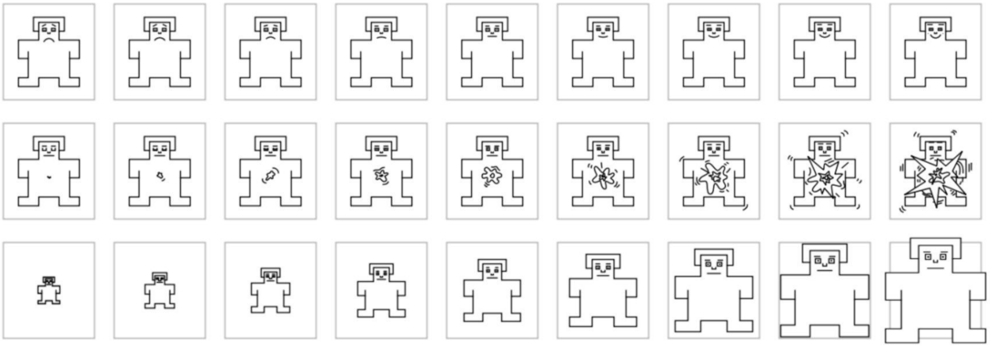


Figure 3.6: Self-Assessment Manikin Questionnaire, three rows representing the valence, arousal, and dominance dimension respectively[2].

The AffectButton (AFB) offers a flexible and dynamic way to collect users' explicit affective feedback (Broekens and Brinkman, 2009, 2013). The AFB is a button like input interface (figure 3.7). In essence, the AFB can be regarded as a navigation tool through a large set of facial expressions. The user can freely move the cursor over the face to change its affective state. Similar to the SAM questionnaire, the AFB returns feedback on the valence, arousal and dominance dimensions. Designed with the intention to be a quick and user-friendly explicit emotion measurement instrument, the reliability and validation of the AFB have been studied on measuring emotional reactions to words, feelings and music (Broekens and Brinkman, 2009; Broekens et al., 2010).

[2] Copyright © 2001-2006, Hans Irtel. Distributed under the MIT License as certified by the Open Source Initiative.

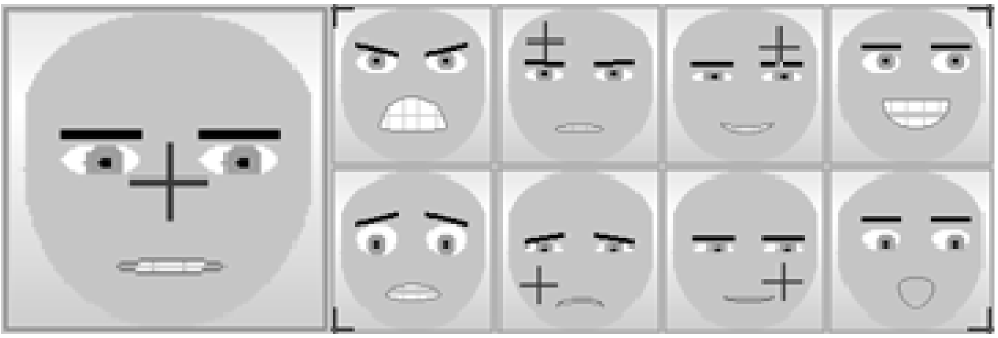


Figure 3.7: AffectButton and its different appearances while moving the cursor (the cross)

3.3.5 Procedure

Prior to the experiment, participants were provided with an information sheet, and the procedure was explained to them. They were then asked to sign an informed consent form. The experiment was setup as a within-subject design, comprising seven conditions with different emotional expressions both in the listening and speaking phase. In each condition, the participants were asked to watch a short clip (around 1 minute) of a conversation about going to conferences between a Chinese virtual lady and a person. In each clip, the virtual human spoke 10 sentences in total, and was silent in between each sentence, listening to her conversational partner talking. The total length of the virtual human's speaking phase was around 15 seconds, and the rest of the 45 seconds was counted as the virtual human's listening phase. The conversation was in Chinese and the participants could hear what the virtual human said during the speaking phase; during the listening phase, there was no sound of the virtual human's conversational partner. The participants were asked to rate the virtual human's emotional state using both SAM and AFB when they finished watching a clip. The order in which the video clips were shown was randomized across the participants.

3.4 Results

The experiment had seven conditions (figure 3.2), with two different measurements and two groups of participants (Chinese and non-Chinese). The data recorded by the SAM questionnaire were integers ranging from 0 to 8, while the data recorded by the AFB were floating-point numbers ranging from -1 to 1. To compare these two measurements, the data were first normalized into z-scores per measurement for each participant across the seven conditions.

The means for the SAM questionnaire and AFB on the valence emotional dimension are shown in figure 3.8. A repeated-measures MANOVA was conducted to test the difference between SAM and AFB scores thereby using condition, type of measurement and cultural background as three independent variables, and the z-scores on valence as dependent variable. The analysis also included all two-way and three-way interactions. The results showed no significant difference between SAM and AFB measurement, $F(1, 22) = 1.30, p = .26$, and also no significant interaction effect.

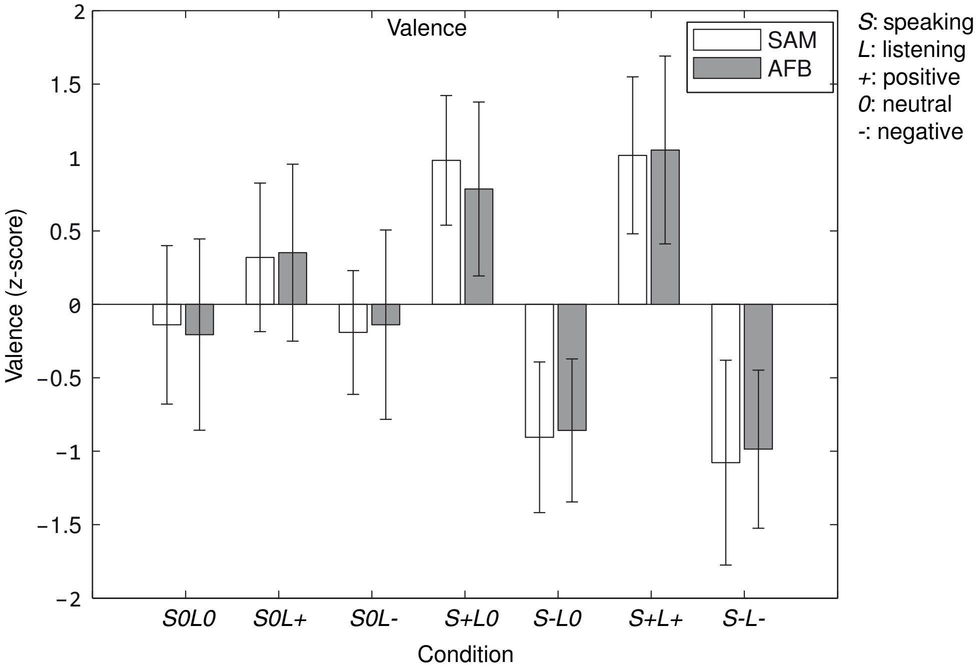


Figure 3.8: Means and standard deviations of SAM and AFB z-scores for the valence dimension for each of the seven experimental conditions.

To test the relationship between these two measurements, a correlation analysis between SAM and AFB scores on the valence dimension was performed. The average scores across all participants for the seven conditions were used. The results showed that SAM and AFB were highly correlated on the valence dimension ($r = 0.995, p < .001$). The valence scores collected by these two measurements could therefore be regarded as consistent. This made it possible to only focus on the average of the SAM and AFB z-scores in the remaining analyses.

3.4.1 Chinese versus non-Chinese

To test the effect of cultural background on the valence rating of the emotional expressions, a mixed MANOVA was conducted using condition as a within-subjects independent variable, cultural background as a between-subjects independent variable, and averaged valence score of both measurements as a dependent variable. The results showed no significant main effect for the cultural background on valence score $F(1, 22) = 1.23, p = .64$, and no significant interaction between cultural background and condition $F(6, 17) = 0.72, p = .28$.

Instead of looking for a difference between participants from different cultural backgrounds, the next step of the analysis focused on similarity in the ratings between these two groups. To examine the relationship between the ratings of Chinese and non-Chinese participants, we performed a correlation analysis based on the means for the seven conditions. The results showed that the scores on valence of the Chinese participants are significantly correlated with those of the non-Chinese participants $r = .98, p < .001$. Although a difference in cultural background was expected, the result showed a high consistency in the evaluation of the emotional state between participants from different cultures. Hence, the results of the two groups of participants were grouped in the rest of the data analyses.

3.4.2 Positive versus Neutral versus Negative Emotional State

Participants were asked to rate seven conditions (i.e., different combinations of a positive, negative and neutral emotional state during the virtual human's speaking and listening phase). A repeated-measures MANOVA was conducted to study the effect of these conditions on averaged valence score of the SAM and AFB z-scores. The results showed a significant effect of condition on the valence rating, $F(6, 18) = 59.50, p < .001$. Next, to run a priori comparisons, paired-sample t-tests were performed using the averaged valence scores of the SAM and AFB z-scores in all the conditions as paired variables. The results are shown in table 3.1.

To test whether the subjective valence score was correlated with the intensity weighted predicted valence values (see chapter 3.3.3 and figure 3.5, and hereafter abbreviated as weighted valence values) for each condition, we calculated the Pearson correlation coefficient between the weighted valence values and the subjective scores averaged over the participants across the seven experimental conditions. This correlation was relatively high, $r = .93, p = .002$. The following step in the analysis was to determine the deviation between the subjective valence scores and their corresponding expected valence value per experimental condition. To do so, we fitted a line through the three data points: `S+L+`, `S0L0`

Table 3.1: Mean, SD and Mean difference of the valence rating of the different conditions.

			Mean Difference / Conditions					
Condition	M	SD	S0L+	S0L-	S+L0	S-L0	S+L+	S-L-
S0L0	-0.17	0.50	-0.51*	-0.008	-1.06*	0.71*	-1.21*	0.86*
S0L+	0.34	0.47		0.500*	-0.55*	1.22*	-0.70*	1.37*
S0L-	-0.17	0.47			-1.05*	0.72*	-1.20*	0.87*
S+L0	0.88	0.39				1.76*	-0.15	1.92*
S-L0	-0.88	0.42					-1.91*	0.15
S+L+	1.03	0.49						2.06*
S-L-	-1.03	0.52						

$H_0 : \mu_1 = \mu_2, ^* p < 0.05.$

and S-L- using least-squares regression. Figure 3.9 shows this line, including the mean subjective valence scores of the remaining four conditions. Deviations of perceived valence from this line (for the lowly negative and positive videos) show to what extent the perceived valence is different from what is expected in case of an equal intensity in valence between the speaking and listening phase (noted as expected valence value in figure 3.9).

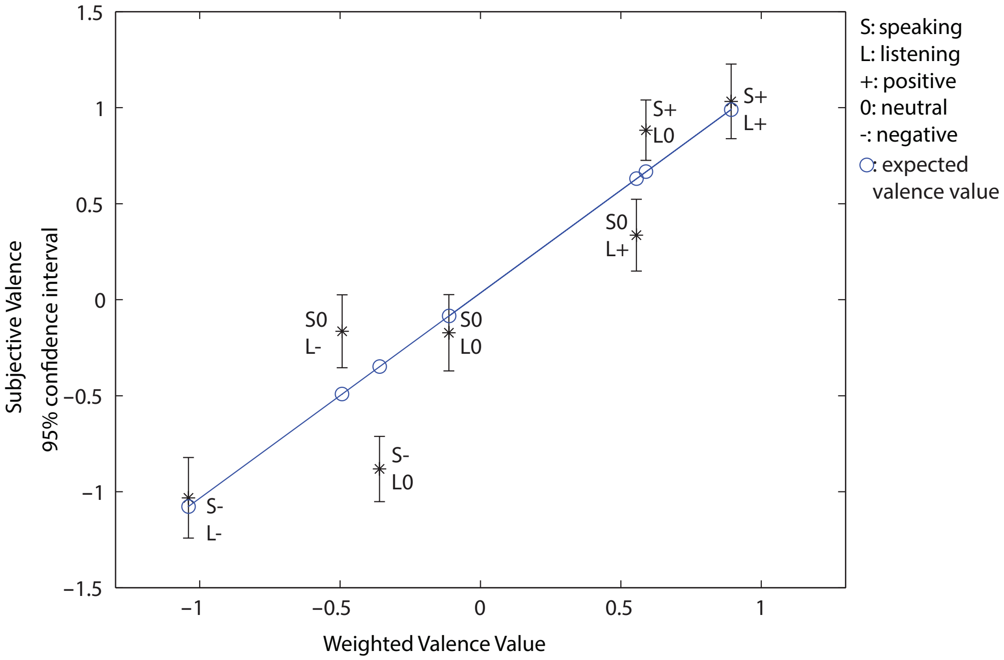


Figure 3.9: The relationship between intensity weighted predicted valence and the averaged subjective valence.

One-sample t-tests revealed that when the virtual human showed neutral listening, both a positive (S+L0) and negative (S-L0) emotional expression during speaking had a more extreme valence than expected, i.e., the subjective score was more positive than the expected valence value in case of the positive emotional expression ($t(23) = 2.69, p = .013$) and more negative than the expected valence value in case of a negative emotional expression during speaking ($t(23) = -6.14, p < .001$). The opposite was seen for the impact of the listening phase. Considering the speaking phase with a neutral emotional expression, the subjective valence score for listening with a positive emotional expression (i.e., the S0L+ condition) was significantly less positive than expected ($t(23) = -3.08, p = .005$). Similarly, the subjective valence score for listening with a negative emotional expression (i.e., the S0L- condition) was significantly less negative than the expected valence value ($t(23) = 3.38, p = .003$).

Moreover, the subjective valence score for the S0L- condition was almost equal ($t(23) = 0.059, p = .95$) to the subjective valence score for the S0L0 condition (i.e., speaking with a neutral emotional expression and listening with a neutral emotional expression). Still, the subjective valence score of the S0L+ condition (i.e., speaking with a neutral emotional expression and listening with a positive emotional expression) was significantly more positive than that for the S0L0 condition ($t(23) = 2.92, p = .008$). A direct comparison of the lowly positive or negative conditions provided a similar pattern. The subjective valence value for the S-L0 condition (i.e., speaking with a negative emotional expression and listening with a neutral emotional expression) was significantly more negative than the subjective valence value for the S0L- condition (i.e., speaking with a neutral emotional expression and listening with a negative emotional expression), $t(23) = 4.97, p < .001$. Similarly, the subjective valance value for the S+L0 condition (i.e., speaking with a positive emotional expression and listening with a neutral emotional expression) was significantly more positive than the subjective valence value for the S0L+ condition (i.e., speaking with a neutral emotional expression and listening with a positive emotional expression), $t(23) = 4.01, p = .001$.

Together these observations imply that people do not perceive much difference between the virtual human showing neutral or negative listening behavior, but they do perceive a difference with the virtual human showing positive listening behavior. In conclusion, all these results support hypothesis 2, stating that the valence of the emotional expression during the listening phase of a conversation is perceived as less impactful compared to the emotional expression during the speaking phase.

Finally, we also compared the more extreme emotional conditions with the $S0L0$ condition. The S+L+ condition ($t(23) = -9.00, p < .001$) or S-L- condition ($t(23) = 5.16, p < .001$) with positive or negative emotional expressions in both the listening and speaking phase, respectively, strongly impact the perceived valence in the expected way.

3.5 Discussion and conclusion

The experiment described in this paper is a human perception study on positive and negative emotions of a virtual human and how cultural background might affect the perception of these emotions. In a sense this study can be seen as a reconfirmation in virtual reality of what is known about human-human interaction in the actual world. Still this is an important validation step as conversations with virtual humans are increasingly used as part of gaming (e.g., Hudlicka and Broekens, 2009), training (e.g., Broekens et al., 2012a), or psychotherapy (e.g., Opris et al., 2012).

The study found that both Chinese and non-Chinese participants could perceive the valence of the virtual human's emotional states and no significant difference between these two groups was found. Instead, the ratings of these two groups were highly correlated. The results show that the valence of the emotional states of the virtual human can be easily recognized by all participants independent of their cultural backgrounds. Hypothesis 1 is therefore not confirmed. On the contrary, our results support the idea of universality of the facial expression of emotion Ekman (1994); Matsumoto (2007), and question the need for tailored made virtual reality applications which target different cultural groups or have multi-cultural users. Still, the results of this study may not be generally applicable to all cultures, since we here only evaluated possible differences in emotion perception between Chinese and non-Chinese people. Further studies are needed to extend our conclusion of universality of emotion perception of virtual human to people with other cultural backgrounds.

In addition, comparing the difference between conditions, it seems that the participants' perception of the valence was more influenced by the emotion of the virtual human while speaking than while listening; and so, this supports Hypothesis 2. Comparing the subjectively perceived valence scores with the expected valence values (figure 3.9), the valence perceived by the participants in the conditions where the listening was neutral, but the speaking performed with a positive or negative emotional expression, was significantly more extreme than what was predicted from equal intensity between speaking and listening. Similarly, the perceived valence was less extreme than the weighted valence value when the speaking was neutral, but the listening performed with a positive or negative emotional expression. This shows the additional influence of verbal communication on valence recognition during a human-virtual human conversation. These findings seem to be in contrast to reports of De Melo et al. (2011), who claim that there is no difference in emotion perception between verbal and non-verbal communication. Their study however used text typing as verbal communication means between human and virtual human, which might explain the different finding. It seems not surprising that the combination of both verbal and non-verbal communication transfers more emotional information than the non-verbal communication only. Furthermore, the influence of the voice can be regarded as content independent because of the high consistency found between the Chinese and non-Chinese participants in this experiment. In other words, the results suggest that affective aspects can be conveyed in the speech even if the language is not understood.

The finding that the perceived valence of the emotion of the virtual human is more intense in the speaking phase than in the listening phase of a conversation may be extended with new research on how to control the level of emotion during these separate phases. Applications such as virtual reality exposure therapy for patients suffering from social phobia may be designed in a way to manipulate the potential phobic stressor using the virtual human's emotional behavior. Further studies may exploit the difference in valence perception between the speaking and listening phase, and explore how to further optimize the persuasive power during these two phases, which may be beneficial for the design of many virtual applications involving human-virtual human conversation. Besides, this study only focuses on how individuals perceive the performance of a virtual human. It is also interesting to test the emotional influence on a human during a human-virtual human conversation. Whether the virtual human's emotion could lead or alter the content of the conversation could be an appealing topic in the persuasive computing area.

Two main conclusions may be drawn from the experiment, but there are also still a number of limitations. First, the virtual human only showed her upper body and no gestures were used to express emotion. However, in recent decades, more insights have become available on body expression (Gross et al., 2010; Kleinsmith and Bianchi-Berthouze, 2013). It would therefore be interesting to examine how our findings would be affected when the virtual human used its full-body to express emotions. Second, the position of the virtual human was fixed in the current study. It would be interesting to test the emotional impact of manipulating the virtual human's position, for example, far away versus nearby Broekens et al. (2012b). Third, the face model of the virtual human we used in this study was generated by FaceGen with the ethnicity parameter set at Southeast Asia. However, no empirical validation was done to confirm the ethnic appearance of the virtual human. Fourth, the study described in this paper only focused on the valence dimension of the emotion, neglecting so far the other two dimensions of emotion, namely arousal and dominance. Including the additional two dimensions would allow to study more complex emotions, for example, fear, surprise, etc. Despite of the limitations, the results of this paper suggest a superior impact on perceiving the virtual human's emotional state during its speaking phase, and a potential independence of the perceived valence of the virtual human's emotion with cultural background. These findings could help designers to focus their attention upon creating and evaluating virtual human with appropriate emotional expressions, which may help to improve the overall experience of virtual environments.

Bibliography

Bailenson, J. N., Yee, N., Merget, D., and Schroeder, R. (2006). The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction. *Presence: Teleoperators and Virtual Environments*, 15(4):359–372.

Barrett, L. F. (2006). Solving the emotion paradox: categorization and the experience of emotion. *Personality and social psychology review : an official journal of the Society for Personality and Social Psychology, Inc*, 10(1):20–46.

Bartneck, C. (2001). Affective expressions of machines. *CHI '01 extended abstracts on Human factors in computing systems*, pages 189–190.

Bates, J. (1994). The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125.

Bradley, M. M. and Lang, P. J. (2007). The International Affective Digitized Sounds (IADS-2): Affective ratings of sounds and instruction manual. *Technical report B-3. University of Florida, Gainesville, Fl.*

Brand, M. (1999). Voice puppetry. *Proceedings of the 26th annual conference on Computer graphics and interactive techniques SIGGRAPH 99*, pages(April):21–28.

Bregler, C. (1997). Video rewrite: Driving visual speech with audio. *Proceedings of SIGGRAPH 97.*, pages 1–8.

Brinkman, W.-P., Hartanto, D., Kang, N., De Vliegher, D., Kampmann, I. L., Morina, N., Emmelkamp, P. M. G., and 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*, pages 1099–1102.

Brinkman, W.-P., Veling, W., Dorrestijn, E., Sandino, G., Vakili, V., and Van der Gaag, M. (2011). Using Virtual Reality to Study Paranoia in Individuals With and Without Psychosis. *Journal of CyberTherapy and Rehabilitation*, 4(2):249–251.

Broekens, J. and Brinkman, W.-P. (2009). Affectbutton: Towards a standard for dynamic affective user feedback. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–8. IEEE.

Broekens, J. and Brinkman, W.-p. (2013). AffectButton: A method for reliable and valid affective self-report. *International Journal of Human-Computer Studies*, 71(6):641–667.

Broekens, J., Harbers, M., Brinkman, W.-P., Jonker, C., Van den Bosch, K., and Meyer, J. J. (2011). Validity of a Virtual Negotiation Training. In *IVA'11 Proceedings of the 11th international conference on Intelligent virtual agents*, pages 435–436. Springer.

Broekens, J., Harbers, M., Brinkman, W.-P., Jonker, C., Van den Bosch, K., and Meyer, J.-J. (2012a). Virtual reality negotiation training increases negotiation knowledge and skill. In *IVA'12 Proceedings of the 12th international conference on Intelligent Virtual Agents*, pages 218–230.

Broekens, J., Pronker, A., and Neuteboom, M. (2010). Real time labeling of affect in music using the AffectButton. In *Proceedings of the 3rd international workshop on Affective interaction in natural environments*, pages 21–26. ACM.

Broekens, J., Qu, C., and Brinkman, W.-P. (2012b). Dynamic Facial Expression of Emotion Made Easy. In *Technical report. Interactive Intelligence, Delft University of Technology*. Technical report. Interactive Intelligence, Delft University of Technology.

Cassell, J., Pelachaud, C., Badler, N. I., Steedman, M., Achorn, B., Becket, T., Doubille, B., Prevost, S., and Stone, M. (1994). Animated Conversation: Rule-based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents. In Huhns, M. N. and Singh, M. P., editors, *Proc of ACM SIGGRAPH*, pages 413–420. ACM Press.

Cassell, J. and Thorisson, K. R. (1999). The Power of a Nod and a Glance: Envelope vs. Emotional Feedback in Animated Conversational Agents. *Applied Artificial Intelligence*, 13:519–538.

Cerezo, E. and Baldassarri, S. (2008). Affective Embodied Conversational Agents for Natural Interaction. In Or, J., editor, *Affective Computing*, chapter 18, pages 329–354. The MIT Press.

Chuang, E. and Bregler, C. (2002). Performance driven facial animation using blendshape interpolation. *Computer Science Technical Report, Stanford University*, 2(2):3.

Church, T. and Katigbak, M. (1998). Language and organisation of Filipino emotion concepts: Comparing emotion concepts and dimensions across cultures. *Cognition & Emotion*, pages 63–92.

Cohen, M. M. and Massaro, D. W. (1993). Modeling coarticulation in synthetic visual speech. In Thalman, N. M. and Thalman, D., editors, *Models and Techniques in Computer Animation*, pages 139–156. Springer-Verlag.

Core, M., Traum, D., Lane, H. C., Swartout, W. R., Marsella, S., Gratch, J., and Van Lent, M. (2006). Teaching negotiation skills through practice and reflection with virtual humans. *Simulation*, 82:685–701.

Cowell, A. J. and Stanney, K. M. (2003). Embodiment and Interaction Guidelines for Designing Credible, Trustworthy Embodied Conversational Agents. In Rist, T., Aylett, R., Ballin, D., and Rickel, J., editors, *4th International Workshop on Intelligent Virtual Agents IVA 2003*, volume 2792 of *Lecture Notes in Computer Science*, pages 301–309. Springer-Verlag.

Darwin, C. (1872). *The Expression of Emotion in Man and Animals*. Number 2 in Project Gutenberg. University of Chicago Press.

De Melo, C., Carnevale, P., and Gratch, J. (2011). The effect of expression of anger and happiness in computer agents on negotiations with humans. *The Tenth International Conference on Autonomous Agents and Multiagent Systems*, pages 2–6.

Dotsch, R. and Wigboldus, D. H. J. (2008). Virtual prejudice. *Journal of Experimental Social Psychology*, 44(4):1194–1198.

Ekman, P. (1994). Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique. *Psychological Bull*, 115:268–287.

Ekman, P. and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 2:124–129.

Ekman, P. and Friesen, W. V. (1978). *Facial action coding system: A technique for the measurement of facial movement*. Consulting Psychologists Press, Stanford University, Palo Alto.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., and Ricci-Bitti, P. E. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53(4):712–717.

Ekman, P., Rolls, E. T., Perrett, D. I., and Ellis, H. D. (1992). Facial expressions of emotion: An old controversy and new findings. *Philosophical Transactions: Biological Sciences*, 335(1273):63–69.

Elfenbein, H. A. (2003). Universals and cultural differences in recognizing emotions. *Current Directions in Psychological*, 12(5):159–164.

Elfenbein, H. A. and Ambady, N. (2002). Is there an in-group advantage in emotion recognition? *Psychological Bulletin*, 128(2):243–249.

Elfenbein, H. A., Beaupre, M., Levesque, M., and Hess, U. (2007). Toward a dialect theory: cultural differences in the expression and recognition of posed facial expressions. *Emotion (Washington, D.C.)*, 7(1):131–146.

Endrass, B., Rehm, M., and Lipi, A. A. (2011). Culture-related differences in aspects of behavior for virtual characters across Germany and Japan. *Proceedings of AAMAS'11*, 2(Section 2):441–448.

Ersotelos, N. and Dong, F. (2008). Building highly realistic facial modeling and animation: a survey. *The Visual Computer*, 24(1):13–30.

Ezzat, T., Geiger, G., and Poggio, T. (2004). Trainable videorealistic speech animation.

Fox, E. (2008). *Emotion Science Cognitive and Neuroscientific Approaches to Understanding Human Emotions.* Palgrave Macmillan.

Gratch, J., Rickel, J., Andre, E., Cassell, J., Petajan, E., and Badler, N. I. (2002). Creating interactive virtual humans: Some assembly required. *Intelligent Systems, IEEE*, 17(4):54–63.

Gross, M. M., Crane, E. A., and Fredrickson, B. L. (2010). Methodology for Assessing Bodily Expression of Emotion. *Journal of Nonverbal Behavior*, 34(4):223–248.

Haring, M., Bee, N., and Andre, E. (2011). Creation and Evaluation of emotion expression with body movement, sound and eye color for humanoid robots. In *RO-MAN, 2011 IEEE*, pages 204–209. IEEE.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organisations across nations.* SAGE Publications, Inc; 2nd edition.

Hudlicka, E. and Broekens, J. (2009). Foundations for modelling emotions in game characters: Modelling emotion effects on cognition. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009.*, pages 1–6.

Irtel, H. (2007). PXLab: The Psychological Experiments Laboratory [online].

Isbister, K. (2006). *Better Game Characters by Design: A Psychological Approach.* CRC Press.

Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., and Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*.

Jan, D., Herrera, D., and Martinovski, B. (2007). A computational model of culture-specific conversational behavior. *Intelligent Virtual Agents*, pages 45–56.

Kahler, K., Haber, J., and Seidel, H.-P. (2001). Geometry-based Muscle Modeling for Facial Animation. In *Proc of Graphics Interface*, pages 37–46. Canadian Information Processing Society.

Keltner, D. and Ekman, P. (2000). Facial expression of emotion. In *Handbook of Emotions, 2nd Edition*, chapter 15, pages 236–249. New York Guilford Publications, Inc.

Kleinsmith, A. and Bianchi-Berthouze, N. (2013). Affective Body Expression Perception and Recognition: A Survey. *IEEE Transactions on Affective Computing*, 4(1):15–33.

Kleinsmith, A., De Silva, P. R., and Bianchi-Berthouze, N. (2006). Cross-cultural differences in recognizing affect from body posture. *Interacting with Computers*, 18(6):1371–1389.

Kulms, P., Kramer, N. C., Gratch, J., and Kang, S.-H. (2011). It's in Their Eyes: A Study on Female and Male Virtual Humans' Gaze. In *IVA'11 Proceedings of the 11th international conference on Intelligent virtual agents*, pages 80–92.

Lance, B. J., Rey, M. D., and Marsella, S. C. (2008). A model of gaze for the purpose of emotional expression in virtual embodied agents. *AAMAS '08 Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems*, 1:12–16.

Lang, P. J. (1995). The emotion probe. Studies of motivation and attention. *American Psychologist*, 50(5):372–385.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1999). International affective picture system (IAPS): Technical manual and affective ratings. Technical report, Gainesville University of Florida, Center for Research in Psychophysiology.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical Report A-8*.

Lee, J. and Marsella, S. C. (2012). Modeling Speaker Behavior: a Comparison of Two Approaches. In *IVA'12 Proceedings of the 12th international conference on Intelligent Virtual Agents*, pages 161–174.

Lee, J., Prendinger, H., Neviarouskaya, A., and Marsella, S. C. (2009). Learning models of speaker head nods with affective information. *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–6.

Lee, Y., Terzopoulos, D., and Walters, K. (1995). Realistic modeling for facial animation. *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques SIGGRAPH 95*, 95(1):55–62.

Link, M., Armsby, P., Hubal, R. C., and Guinn, C. I. (2006). Accessibility and acceptance of responsive virtual human technology as a survey interviewer training tool. *Computers in Human Behavior*, 22(3):412–426.

Litwinowicz, P. and Williams, L. (1994). Animating images with drawings. *Proceedings of the 21st annual conference on Computer graphics and interactive techniques SIGGRAPH 94*, 28(Annual Conference Series):409–412.

MacDorman, K. F., Coram, J. A., Ho, C.-C., and Patel, H. (2010). Gender differences in the impact of presentational factors in human character animation on decisions in ethical dilemmas. *Presence: Teleoperators and Virtual Environments*, 19(3):213–229.

Matsumoto, D. (2002). Methodological requirements to test a possible in-group advantage in judging emotions across cultures: Comment on Elfenbein and Ambady (2002) and evidence. *Psychological Bulletin*, 128(2):236–242.

Matsumoto, D. (2007). Emotion judgments do not differ as a function of perceived nationality. *International Journal of Psychology*, 42(3):207–214.

Mehrabian, A. and Russell, J. A. (1974). *An Approach to Environmental Psychology*. MIT Press, Cambridge, MA, USA; London, UK.

Moridis, C. N. and Economides, A. A. (2012). Affective Learning: Empathetic Agents with Emotional Facial and Tone of Voice Expressions. *IEEE Transactions on Affective Computing*, 99.

Morris, J. D. (1995). Observations : SAM The Self-Assessment Manikin An Efficient Cross-Cultural Measurement Of Emotional Response. *Journal of Advertising Research*, 35(6):63–68.

Ochsman, R. B. and Chapanis, A. (1974). The effects of 10 communication modes on the behavior of teams during co-operative problem-solving. *International Journal of ManMachine Studies*, 6(5):579–619.

Opris, D., Pintea, S., Garcia-Palacios, A., Botella, C. M., Szamoskozi, S., and David, D. (2012). Virtual reality exposure therapy in anxiety disorders: a quantitative meta-analysis. *Depression and Anxiety*, 29(2):85–93.

Parke, F. I. (1972). Computer generated animation of faces. *Proceedings of the ACM Annual Conference*, 1:451–457.

Parke, F. I. (1974). A parametric model for human faces. *The University of Utah, Doctoral Dissertation.*

Petrushin, V. (1999). Emotion in speech: Recognition and application to call centers. *Artificial Neu. Net. In Engr.(ANNIE'99).*

Picard, R. W. (1998). *Toward agents that recognize emotion.* MIT Media Laboratory, Perceptual Computing Section.

Picard, R. W., Vyzas, E., and Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(10):1175–1191.

Platt, S. M. and Badler, N. I. (1981). Animating facial expressions. *ACM SIGGRAPH Computer Graphics*, 15(3):245–252.

Qiu, L. and Benbasat, I. (2005). Online consumer trust and live help interfaces: The effects of text-to-speech voice and three-dimensional avatars. *International Journal of Human-Computer Interaction*, 19(1):37–41.

Reeves, B. and Nass, C. (1996). *The Media Equation.* Cambridge University Press.

Rizzo, A. A., Lange, B., Buckwalter, J. G., Forbell, E., Kim, J., Sagae, K., Williams, J., Rothbaum, B. O., Difede, J., Reger, G., Parsons, T. D., and Kenny, P. G. (2011). An intelligent virtual human system for providing healthcare information and support. *Stud Health Technol Inform*, 163:503–509.

Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological Bulletin*, 110(3):426–450.

Ruttkay, Z. and Pelachaud, C. (2005). *From Brows to Trust: Evaluating Embodied Conversational Agents.* Springer.

Scherer, K. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1-2):227–256.

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice : Official Journal of the Voice Foundation*, 9(3):235–48.

Schiano, D. J., Ehrlich, S. M., Rahardja, K., and Sheridan, K. (2000). Face to interface: facial affect in (hu)man and machine. In *Proceedings of ACM CHI 2000*, pages 193–200. ACM.

Schroder, M. (2004). *Speech and Emotion Research: an overview of research frameworks and a dimensional approach to emotional speech synthesis.* PhD thesis, Saarland University.

Slater, M., Pertaub, D.-P., and Steed, A. (1999). Public speaking in virtual reality: facing an audience of avatars. *IEEE Computer Graphics and Applications*, 19(2):6–9.

Tomkins, S. S. (1962). *Affect, Imagery, Consciousness: Vol 1. The Positive Affects.* New York: Springer.

Tomkins, S. S. (1963). *Affect, Imagery, Consciousness: Vol 2. The Negative Affects.* New York: Springer.

Tsapatsoulis, N., Raouzaiou, A., Kollias, S., Cowie, R., and Douglas-Cowie, E. (2002). Emotion Recognition and Synthesis Based on MPEG-4 FAPs. In *MPEG-4 facial animation the standard implementations applications*, chapter 9. Wiley, Hillsdale.

Waters, K. (1987). A Muscle Model for Animating Three-Dimensional Facial Expression. *Comput Graph SIGGRAPH Proc*, 21(4):17–24.

Wierzbicka, A. (1995). *Emotions across languages and cultures: Diversity and universals.* Cambridge University Press; 1st edition.

Wong, J. W.-E. and McGee, K. (2012). Frown More, Talk More: Effects of Facial Expressions in Establishing Conversational Rapport with Virtual Agents. In *IVA'12 Proceedings of the 12th international conference on Intelligent Virtual Agents*, pages 419–425.

Yun, C., Deng, Z., and Hiscock, M. (2009). Can local avatars satisfy a global audience? A case study of high-fidelity 3D facial avatar animation in subject identification and emotion perception by US and international groups. *Computers in Entertainment*, 7(2):1.

Zhang, Q., Liu, Z., Guo, B., and Shum, H. (2006). Geometry-driven photorealistic facial expression synthesis.

Chapter 4

The Virtual Dialog Partner II

The effect of synthetic emotions:

an empirical study on human response to a emotional virtual dialog partner

This chapter has been published as:
Chao Qu, Willem-Paul Brinkman, Yun Ling, Pascal Wiggers, Ingrid Heynderickx (2014). Conversations with a virtual human: Synthetic emotions and human responses. *Computer in Human Behaviors*, 34, 58-68.

To test whether synthetic emotions expressed by a virtual human elicit positive or negative emotions in a human conversation partner and affect satisfaction towards the conversation, an experiment was conducted where the emotions of a virtual human were manipulated during both the listening and speaking phase of the dialog. Twenty-four participants were recruited and were asked to have a real conversation with the virtual human on six different topics. For each topic the virtual human's emotions in the listening and speaking phase were different, including positive, neutral and negative emotions. The results support our hypotheses that (1) negative compared to positive synthetic emotions expressed by a virtual human can elicit a more negative emotional state in a human conversation partner, (2) synthetic emotions expressed in the speaking phase have more effect on a human conversation partner than emotions expressed in the listening phase, (3) humans with less speaking confidence also experience a conversation with a virtual human as less positive, and (4) random positive or negative emotions of a virtual human have a negative effect on the satisfaction with the conversation. These findings have practical implications for the treatment of social anxiety as they allow therapists to control the anxiety evoking stimuli, i.e., the expressed emotion of a virtual human in a virtual reality exposure environment of a simulated conversation. In addition, these findings may be useful to other virtual applications that include conversations with a virtual human.

4.1 Introduction

Humans are social creatures for which conversations with others are an essential part of their everyday life. These conversations allow them to influence each other's behaviour, attitudes and emotions. Conversations are part of complex social interactions, such as learning, negotiation, and coordination. Not surprisingly, people strive to become more comfortable and skilled in conducting conversations. With the introduction of virtual reality and virtual humans, people can experience conversations in a controlled simulated environment, for example, to practice various conversation skills including negotiation (Broekens et al., 2012a; Core et al., 2006), communication (Lok, 2006), interview (Link et al., 2006), leadership (Swartout, 2006), and decision making (Wandner et al., 2013). Virtual reality has also been suggested as a treatment environment for individuals with social anxiety, who fear social interaction such as casual or formal conversation settings (Anderson et al., 2004, 2001; Krijn et al., 2004b; Szegedy-Maszak, 2004). The findings of using virtual reality exposure therapy (VRET) for other types of anxiety disorders, e.g., fear of flying or fear of heights, are encouraging as meta-studies (Gregg and Tarrier, 2007; Opris et al., 2012; Parsons and Rizzo, 2008; Powers and Emmelkamp, 2008) indicate that virtual reality exposure is as effective as in vivo exposure, the latter being the golden standard for anxiety disorder treatment.

A key benefit of VRET is the therapist's ability to control the feared stimulus. This is important, as patients need to be gradually exposed, starting with the least feared stimuli, which is then gradually increased to more feared stimuli. In the case of social phobia, this is often implemented as switching between different social scenes, such as buying items in a shop, having a blind date, or speaking in public (Brinkman et al., 2008; Klinger et al., 2004). Emmelkamp (2013), however, suggests that variation within a scene should also be possible in the treatment of social anxiety. VRET systems for the treatment of other anxiety disorders do already provide this. For example, for fear of flying, the therapist can change the weather the airplane is flying through, show safety instructions on the seat's build-in monitor, or let the pilot make an announcement to fasten the seatbelts or to expect turbulence (Brinkman et al., 2010; Gunawan et al., 2004). In treating patients with fear of height, the therapist can choose the vertical location of a patient on for example a virtual staircase, or move the patient closer or further away from the edge of a balcony (Krijn et al., 2004a). For patients with social anxiety, the therapist also needs access to these controls (Clark and Beck, 2011) and needs more flexibility (Lanyi et al., 2011). One potential way of doing this, for social anxiety, is to allow the therapist to control the emotions expressed by the virtual human in a conversation. This would build on recent progress to engage humans in an actual natural verbal conversation with a virtual human (Brinkman et al., 2012; Kwon et al., 2009; Ter Heijden and Brinkman, 2011).

This paper, therefore, studies dialog manipulations that allow therapists to control the fear stimuli that induce different levels of anxiety in social phobic patients. By controlling non-verbal behaviour, such as facial expression and head movement, and verbal behaviour such as voice intonation, the therapists can control the emotions expressed by the virtual human in the dialog.

4.2 Hypotheses

Two decades ago, Reeves and Nass (1996) made a compelling case about the similarity in the way humans response to computers and the way they respond to other humans. Giving a computer agent a human shape can make the interaction with individuals more positive as Yee et al. (2007) found in their meta-analysis. On the other hand, virtual humans can also elicit anxiety in individuals not only by their high level of appearance realism (Kwon et al., 2009) but also by their non-verbal behaviour (James et al., 2003). If virtual humans are capable of having a natural, effective and expressive interaction with people, they can be used in a variety of applications, such as VRET for patients with social anxiety. Our current study is set up around four hypotheses that focus on the effect of synthetic emotions, either being positive, negative, neutral or random, the difference between these emotions expressed when a virtual human is talking or listening, and the difference in response between individuals with low or higher level of speaking confidence. We measure the degree of satisfaction people obtain from a conversation with the virtual human. When considering a conversation as an exchange of questions and answers, satisfaction is defined as “the feeling the user got during the question phase and how the user experienced the answers and attention from the virtual human” (Ter Heijden and Brinkman, 2011). Besides satisfaction, we also measure how the emotions expressed by the virtual human affect the emotional state of an individual. For the formulation of the hypotheses, we specifically focus on the valence dimension of the three-dimensional Valence - Arousal - Dominance Emotion Model (Schlosberg, 1941; Schroder, 2004).

4.2.1 Positive Emotions versus Negative Emotions

Affective feedback plays a key role in a conversation. It may cause defensive or supportive listener’s response (Gibb, 1961). Interestingly, similar effects are reported for virtual worlds. For example, Pertaub et al. (2002) exposed individuals as a speaker to a neutral, positive and negative virtual audience, and found that the audience’s attitude affected the user’s sense of satisfaction. Several researches have also studied the impact of positive behaviour of a virtual human on actual humans. De Melo et al. (2012) found that people disliked negotiating with angry virtual humans and tended to treat them as uncooperative and dominant. At the more positive side, Maldonado et al. (2005) found that positive emotions expressed by a co-learner enhanced student's learning gains and enjoyment, even if the co-learner simply existed of a set of photos of human facial expressions. Also Burleson and Picard (2007) showed that systems with a virtual character that provided affective support reduced frustration of less confident users. All these studies show that virtual humans that express emotions may also affect an individual. Therapists may use this; for example, at an initial stage of an exposure therapy they may use virtual humans expressing positive emotions to limit the amount of anxiety they want to elicit in a patient. Later on in the exposure they may let the virtual human express negative emotions to again elicit anxiety as the anxiety provoking element of having a conversation with a positive virtual human has worn off. Being able to do this would be beneficial for applications such as VRET. Evidence in the literature supports the idea that positive and negative emotions can be elicited in a conversation with a virtual human, but this evidence is basically indirect in the sense that the literature mainly focused on one-way conversations where a single virtual human or audience listened to a human (Ling et al., 2012; Pertaub et al., 2002; Wong and McGee, 2012) or where a virtual human speak to a human (Baylor et al., 2003; Konstantinidis et al., 2009; Qiu and Benbasat, 2005). Here, we systematically examine the effect of emotion expression of a virtual human on its conversational partner in a two-way free-speech dialog. In the context of social anxiety, negative or positive emotion expression refers to expressions of the virtual human from which human conversation partners could deduce that they are negatively or positively evaluated by the virtual human. Thus, the effect of positive and negative emotions of virtual human towards human conversation partner leads to the first hypothesis.

Hypothesis 1: Compared to negative emotions expressed by a virtual human in a conversation with a person, positive emotions expressed by the virtual human result in a more positive emotional state in the person, and also in more satisfaction towards the conversation.

4.2.2 Emotions during Speaking versus Listening

When persons are engaged in a human-human conversation, their behaviour can be separated into two phases: a listening phase and a speaking phase. In the listening phase, emotions are mainly expressed by non-verbal behaviour such as facial expressions. In the speaking phase, non-verbal behaviour is extended with a very dominant verbal component, e.g., by voice intonation. In a natural conversation, these phases may be almost unnoticeably intertwined (Adler, 1997). In a conversation with a virtual human, on the other hand, both phases have been mainly studied separately, focusing on the most critical phase for a specific application. For example, Brinkman et al. (2011) manipulated the emotions expressed by virtual humans when they were speaking with a person in a cloth shop, and as such, varied the amount of stress evoking elements as part of an aggression management environment. They found that when the virtual human was talking aggressively, their participants had higher physiological arousal as compared to the condition where the virtual human was talking passively. Likewise, Konstantinidis et al. (2009) used a talking virtual character that was able to express emotions in an educational environment for autistic children, and found that autistic children were able to recognize the virtual character's mental and emotional state provided by facial expressions, and thus the virtual character advanced the educational process. Other studies focused mainly on the effect of emotions in the listening phase. For example, Wong and McGee (2012) asked their participants to tell stories to an emotional agent and found that the agent's inappropriate emotional feedback such as an incongruous emotional reaction increased story length compared to the agent's appropriate emotional feedback such as a smile or a surprised expression as relevant to the story. Another prominent listening example is a virtual audience created to simulate a public speaking scenario as done by Pertaub et al. (2002). They found that a negative audience elicited a significantly higher level of anxiety in human speakers compared to the neutral and positive audiences. Interestingly, in principle therapists can control both phases of a conversation. Still, when simulated, we need to understand the intensity of the effect raised in patients during both phases. This effect may be unequal since in the speaking phase emotions may be expressed verbally as well, whereas emotions may only be expressed non-verbally in the listening phase. This, therefore, leads to the second hypothesis.

Hypothesis 2: An individual's negative or positive emotion in a conversation with a virtual human and the satisfaction towards this conversation are more affected by the emotions expressed by the virtual human in the speaking phase than in the listening phase.

4.2.3 Low Anxiety Group versus High Anxiety Group

If the dialog manipulations suggested previously have any relevance for the treatment of patients with a social anxiety disorder, these individuals should response more intensely to them. Powers et al. (2013) recently showed that a conversation with a virtual human in virtual reality could indeed elicit anxiety, even more than a similar conversation with an actual person. More specifically, Slater et al. (2006) were able to show that people with a lower speaking confidence were more influenced by the emotions of a virtual human in a public speaking scenario than people with a higher speaking confidence. A follow up study (Pan et al., 2012) also found that this group of people reported a greater sense of being disturbed when the surrounding virtual humans looked towards them. The third hypothesis therefore addresses this difference between people with a low and high speaking confidence.

Hypothesis 3: Compared to individuals with a high degree of speaking confidence, individuals with a low degree of speaking confidence obtain less satisfaction from a conversation with a virtual human, and have a more negative emotional state during the conversation.

4.2.4 Random Emotions versus Neutral Emotions

When implementing synthetic emotions in a simulated conversation, a key question is how much attention one should pay to the consistency of the expressed emotions. With other words, would the conversation experience already improve if the virtual human expresses, even inconsistently, different emotions, instead of having a consistent neutral emotional expression? A related question is what would be the effect if a therapist would often change the parameter settings between positive and negative emotions during a conversation? Switching too often would create inconsistency in the expressed emotions. Human conflict theorists argue that emotion inconsistency creates a sense of unpredictability (Schelling, 1981) and gives observers a sense of uneasiness (Morris, 2002a). People with unpredictable emotion expressions, such as alternating expressing anger and happiness, could cause their negotiation opponents to feel less in control (Sinaceur et al., 2013) and to make greater concessions (Van Kleef and De Dreu, 2010). This therefore leads to the fourth and the final hypothesis.

Hypothesis 4: Compared to neutral emotions expressed continuously by a virtual human in a conversation, positive and negative emotions expressed randomly by a virtual human result in less satisfaction towards the conversation.

4.3 Method

A within-subjects experiment with six conditions (see Table 4.1) was setup to test the four hypotheses. Specifically, for testing the second hypothesis, the emotion expression in the speaking phase (`S`) and listening phase (`L`) was separately controlled. This makes it possible for the virtual human to express positive emotion (indicated by `+`) while talking but negative emotion (`-`) while listening, or vice versa. In order to test hypotheses 1 and 2, a 2-by-2 within-

Table 4.1: Six experimental conditions.

Condition	Listening phase	Speaking phase
L+S+	Positive	Positive
L+S-	Positive	Negative
L-S+	Negative	Positive
L-S-	Negative	Negative
L0S0	Neutral	Neutral
LrSr	Random	Random

subjects design with four conditions (i.e., L+S+, L+S-, L-S+, L-S-) was created. So, for example in the L+S- condition, the virtual human was positive when listening and negative when speaking.

In addition, to test the fourth hypothesis, two other conditions were also created: a neutral (indicated by 0 in Table 4.1) condition and a random (indicated by r in Table 4.1) condition. In the neutral condition, the virtual human was completely neutral both in the speaking and listening phase. In the random condition, the virtual human showed either positive or negative emotions in completely random order both in the speaking and the listening phase. So, the emotion expressions varied between the speaking and listening phase, and from sentence to sentence.

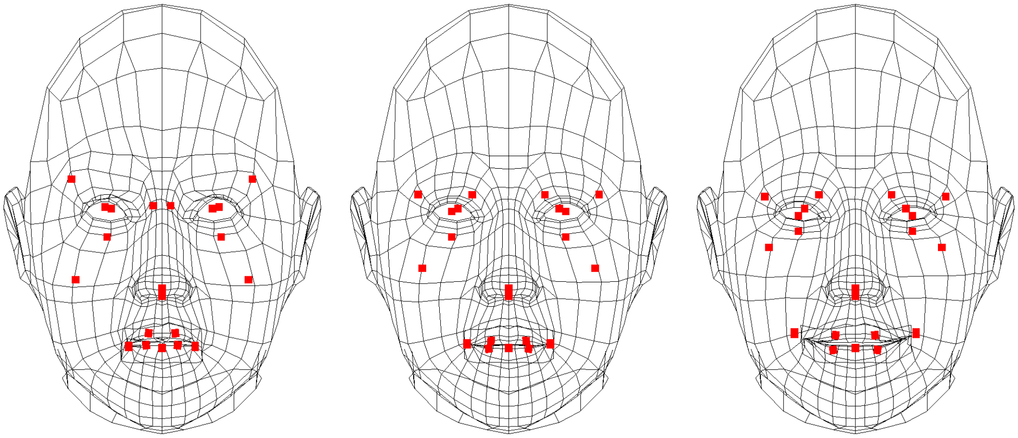
4.3.1 Participants

Twenty-four Chinese (11 female and 13 male) students from the Delft University of Technology participated in the experiment. Their age ranged from 24 to 30 years with the mean being 26.4 ($SD = 1.6$) years. All participants were native speakers of mandarin Chinese and they were all naive with respect to the four hypotheses until they finished the experiment. Written informed consent was obtained from all participants prior to the experiment. All participants received a small gift for their contribution. The experiment was approved by the Delft University of Technology Human Research Ethics Committee, and was done in accordance to local ethical customs.

4.3.2 Apparatus

Cowell and Stanney (2003) found that people generally prefer to interact with a youthful character matching their ethnicity, but they did not find a significant preference for the gender of the character. Furthermore, Kulms et al. (2011) showed that actual behaviour is more important than gender stereotypes for the evaluation of the interaction. Therefore, a Chinese female virtual character aged around 25 was specially created for this study.

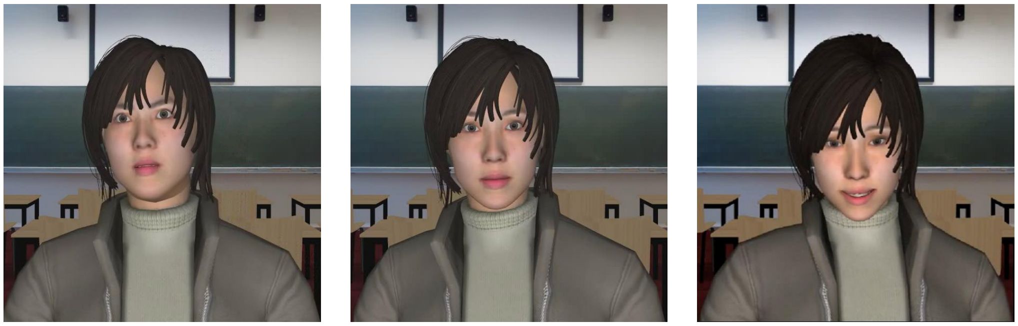
The model of the Chinese lady was created with FaceGen and 3Ds MAX. Several factors, which were considered to contribute to her emotional expression during the conversation, were manipulated: her facial expressions, her head movements, her eye movements and her voice intonation. A repeated facial expression animation method was used to generate facial expressions. This method rigged the face mesh with 22 action units and 18 features (Gratch et al., 2002), and each feature had an anchor point attached to a set of vertices of the face as control points. A model of dynamics that could control the intensity of the expression, the onset, peak and decay was defined. This model gave the virtual human the ability to show any intensity and any combination of the six basic Ekman facial expressions (Ekman and Friesen, 1978). By setting the values for the three emotion dimensions (i.e., valence, arousal and dominance), and the expression duration, any emotion could be expressed (Broekens et al., 2012b). Figure 4.1 shows the virtual human expressing emotions from neutral (b) to negative (a) or positive (c).



(a) angry (b) neutral (c) happy

Figure 4.1: Emotions expressed by moving action units (i.e., small squares) attached to a face mesh

During the listening phase, the virtual lady showed a happy facial expression in the positive condition. She also nodded her head once in a while to agree with what the participant said. Her eyes looked away only occasionally, but most of the time, she looked at the participant (Figure 4.2(c)). In the negative condition, on the other hand, she had an angry facial expression and looked away most of the time. She showed only limited interest in her conversation partner - the participant (Figure 4.2(a)). The intensity of both the positive and negative emotional expressions was evaluated in a previous study (Broekens et al., 2012b) to ensure that they both could be identified by individuals. For the neutral condition, a neutral facial expression[1] was used and the lady kept looking at the participants with some slight eye and head movements (Figure 4.2(b)). In the random condition, the Chinese lady had an unstable emotional expression. At one moment in time, she appeared positive, but one moment later when she finished her sentence and started listening she could become negative. The chance of her being positive or negative was 50% - 50%, and she would only change her behaviour at the beginning of every speaking or listening phase.



(a) Negative: angry facial expression, only looking at her conversation partner at the beginning, gradually losing interest and starting to look around.

(b) Neutral: neutral facial expression while constantly looking at her conversation partner with some slight eye movements.

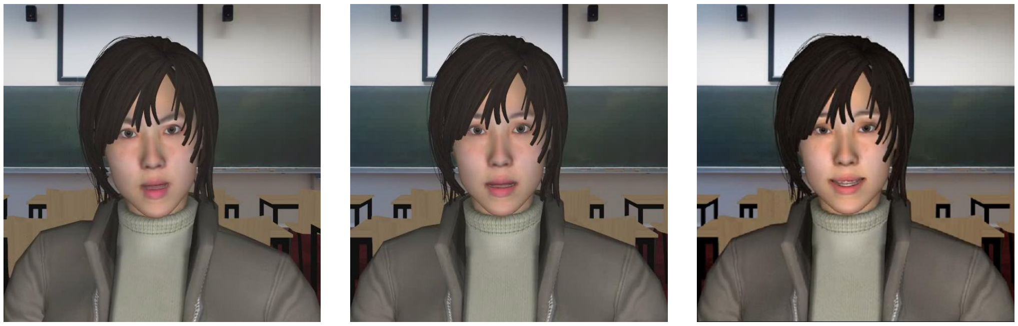
(c) Positive: happy facial expression while constantly looking at her conversation partner, showing some slight eye movements, and occasionally nodding her head.

Figure 4.2: Different emotional states of the virtual human in her listening phase

During the entire speaking phase, the virtual lady looked directly at the participants. An angry facial expression was shown in the negative condition (Figure 4.3(a)) and a happy facial expression was shown in the positive condition (Figure 4.3(c)). In addition, negative / positive voice intonation was added to the corresponding conditions. For the neutral condition, neutral voice intonation was used instead and the lady showed a neutral facial expression (Figure 4.3(b)). Again, the random condition existed of the combination of positive and negative emotions, controlled by a random coefficient.

Since the participants were asked to have a real question and answer session with the virtual lady, the verbal behaviour of the virtual lady was manipulated by an experimenter located behind a shielding screen. The dialog tool Editor3 (Ter Heijden and Brinkman, 2011; Ter Heijden et al., 2010) was used to create six

[1] The default facial expression generated by FaceGen with the parameters for the six basic emotion expressions set to zero and any other morph modifiers removed.



(a) Negative: angry facial expression while looking at her conversation partner, and speaking with a negative voice intonation.

(b) Neutral: neutral facial expression while constantly looking at her conversation partner, and speaking with a neutral voice intonation.

(c) Positive: happy facial expression while constantly looking at her conversation partner, and speaking with a positive voice intonation

Figure 4.3: Different emotional states of the virtual human in her speaking phase

dialogs on the following topics: research project, food, movie, China, travelling, and living in the Netherlands. Each dialog consisted out of ten main questions and on average two follow-up questions for each main question. Based on what the participant said during the conversation, the experimenter would select an appropriate voice recorded response for the virtual lady from a set of on average three responses. A conversation lasted on average 411 seconds ($SD = 137$).

Figure 4.4 shows the setup of the experiment. To make the experiment double blind, the participants wore an earphone to listen to the virtual lady. This way, the experimenter could neither see the emotional expression of the virtual lady nor hear her voice intonation. This ensured he was unaware of the experimental condition.

4.3.3 Validation of the Stimuli

The voice of the virtual lady used in this experiment was recorded in Chinese by a Chinese linguistics student. Each single sentence was recorded three times. The content was each time the same, but the intonation was different: once neutral, once positive and once negative. To validate the recordings, a small preliminary study with 6 Chinese participants (3 male and 3 female) with an average age of 27 ($SD = 0.5$) years was conducted. These participants were all students from Delft University of Technology and they were all native speakers of mandarin Chinese. To avoid a possible learning effect, these participants did not participate in the main experiment. They were asked to rate the valence of the recorded voice on a scale from 1 (negative) to 9 (positive). As the dependent variable deviated from normality, non-parametric analyses were conducted. The

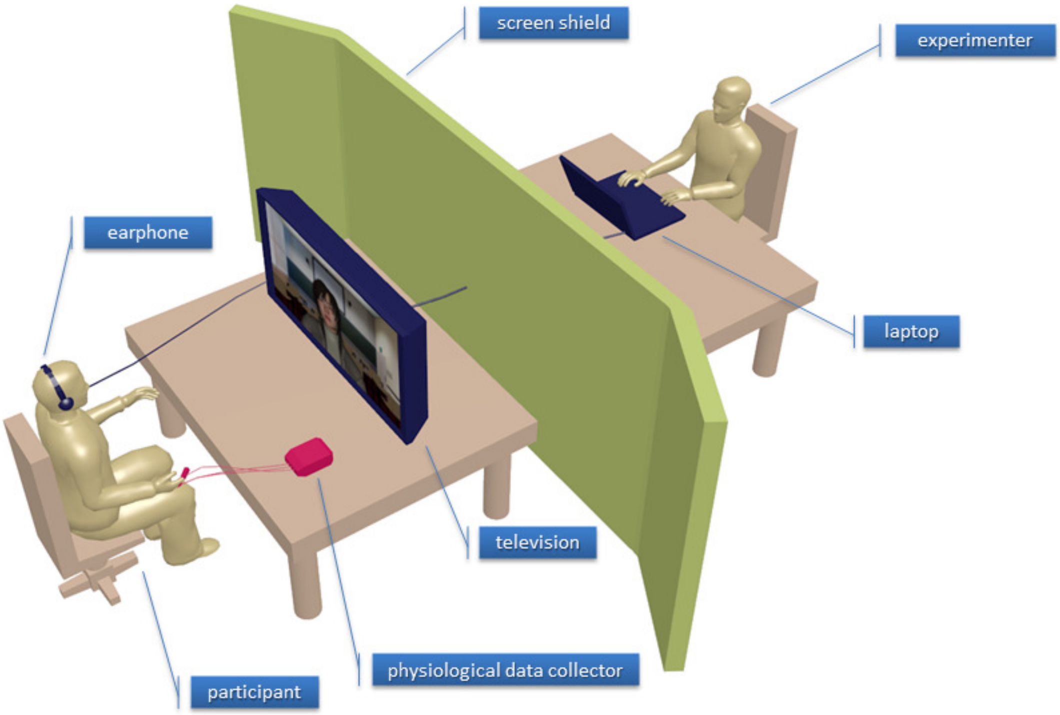


Figure 4.4: Experimental Setup

result of a Friedman test showed that the emotion in the recorded voice was indeed perceived as intended ($\chi^2(2, N = 6) = 11.57, p = .003$). The result of Wilcoxon Signed Ranks tests showed that the positive voice received a significantly higher valence rating than the neutral voice ($z = 2.03, p = .042$), and the negative voice ($z = 2.21, p = .027$). Furthermore, the negative voice received a significantly lower valence rating than the neutral voice ($z = 2.21, p = .027$). The medians and interquartile ranges (in brackets) of the scores on the positive, neutral and negative voice were 8.5 (2.0), 5.5 (6.0) and 1.5 (1.0) respectively.

For testing Hypothesis 2, a fair comparison between the listening and speaking phase was needed, which meant that the intensity of the non-verbal communication in both phases should be similar. For example, the virtual lady's facial and body expressions in the negative speaking phase should have a similar valence impact as in the negative listening phase. To test this, another small preliminary study was conducted using sound exclusive videos of the virtual lady during a conversation. Twelve participants, 5 male and 7 female with an average age of 27 years ($SD = 1.8$) were presented simultaneously with two video clips of the virtual lady, one of the listening and one of the speaking phase. These participants were all students from Delft University of Technology. Half of the participants were Chinese, and all these participants again did not participate in the main experiment. The participants were asked to rate how easily they could see the difference between the two videos on a scale from very easy (0) to very difficult (100). The participants were explicitly asked not to rate the valence, but only the easiness with which differences were perceived, representing the intensity of the emotion. The participants were asked to rate 12 pairs in total (S-L0/S0L0, S0L-/S0L0, S+L0/S0L0, S0L+/S0L0, S-L-/S+L+, S-L-/S0L0, S+L+/S0L0, S0L0/S0L0, S+L0/S+L0, S-L0/S-L0, S0L+/S0L+, S0L-/S0L-). Before they rated the pairs, the participants were shown all the possible behaviours of the virtual human so that they could establish an overall frame of reference.

As all the dependent variables were normally distributed, a parametric test, i.e., MANOVA with repeated measures was conducted with the valence direction and the phase (speaking versus listening) as independent variables. The analysis only used the ratings for the only positive speaking (S+L0/S0L0) and only positive listening ($S0L+/S0L0$) pairs, and the ratings for the only negative (S-L0/S0L0) speaking and only negative listening ($S0L-/S0L0$) pairs. The analysis revealed that the positive videos ($M = 28, SD = 16$) were rated significantly ($F(1, 11) = 16.91, p = .002$) easier to be distinguished than the negative videos ($M = 59, SD = 24$) from the neutral reference video ($M = 85, SD = 16$). But no significant difference was found between the listening and speaking phase ($F(1, 11) = 0.14, p = .71$), and also no significant two-way interaction effect was found ($F(1, 11) = 0.44, p = .52$). The results showed that compared to the neutral reference video, the positive or negative differences from neutral in the listing or speaking phase were equally distinguishable, and so, the intensity of the non-verbal communication was similar in the listening and speaking phase.

4.3.4 Measurements

Personal Report of Confidence as a Speaker

The Personal Report of Confidence as a Speaker (PRCS) questionnaire (Paul, 1966) was used as a screening test for everyday experienced fear of speaking. It is a self-report questionnaire that assesses the behavioural and cognitive response to public speaking. The PRCS questionnaire recorded whether participants agreed or disagreed on 30 statements, for example *"I dislike to using my body and voice expressively."* The PRCS index was scored by counting the number of answers indicating anxiety. The PRCS index ranges from 0 to 30. Daly (1978) reported strong correlations between the PRCS index and other social phobia measures. Furthermore, Phillips et al. (1997) showed that the PRCS index did not differ across age and gender.

Dialog Satisfaction

The Dialog Experience Questionnaire (DEQ) (Ter Heijden and Brinkman, 2011) was used to measure the participant's satisfaction towards the conversation with the virtual lady. The DEQ has four flow sub-dimensions (i.e., dialog speed, interruption, correctness locally and correctness globally) and two interaction sub-dimensions (i.e., involvement and discussion satisfaction). In the analysis only the mean of the five items addressing the sub-dimension discussion satisfaction were considered. As a consequence, the score ranged from -3 to 3.

Self-Assessment Manikin questionnaire

The Self-Assessment Manikin Questionnaire (SAM) (Lang, 1995) was included to subjectively measure the three emotion dimensions, i.e., valence, arousal and dominance. Various studies showed that the SAM questionnaire accurately measured emotional reactions to imagery (Lang et al., 1999; Morris, 1995), sounds (Bradley and Lang, 2007), robot gesture expression (Haring et al., 2011), etc. The SAM questionnaire consists of a series of manikin figures to judge the affective quality and represents the intensity value of the three dimensions of emotion (Lang, 1995). The first row of SAM manikin figures ranges from unhappy (1) to happy (9) on the valence dimension. The second row represents the arousal dimension, ranging from relaxed (1) to excited (9). The last row ranges from dominated (1) to controlling (9), representing the dominance dimension. After being explained the meaning of each dimension, participants selected one of the nine figures on each row to express their feelings during the conversation. The manikin figures were taken from the PXLab (Irtel, 2007).

Presence questionnaire

Participants were asked to complete the Igroup Presence Questionnaire (IPQ) (Schubert et al., 2001) to measure their experienced presence during the conversation. IPQ comprises out of 14 items rated on a seven-point Likert Scale. The scores on the 14 IPQ items are mapped onto three subscales, namely Involvement (i.e., the awareness devoted to the virtual environment), Spatial Presence (i.e., the relation between the virtual environment and the physical real world), and Realism (i.e., the sense of reality attributed to the virtual environment). It also contains one item that assesses the general feeling of being in the virtual environment. The total score of IPQ was used in the data analysis to test whether the level of presence was sufficient to evoke an emotional response in the participants. The total score of IPQ ranged from 0 to 84.

Dialog length

Gratch and Okhmatovskaia (2006) found that people talked longer to a responsive than to an unresponsive virtual human. Also Wong and McGee (2012) showed that people talked longer to a virtual human that listened with a slight frown or responded to the speaker's facial expression with sadness or puzzlement than to a virtual listener that showed a small smile and mirrored the positive emotional expressions of the human speaker. Speaking time has also been suggested as a reliable behavioural measure to assess performance anxiety (Beidel et al., 1989). As such, in an impromptu speech task, patients are asked to give a speech, and the length of the speech is taken as reversed indicator of avoidance behaviour. Therefore, in this experiment the total time a participant talked during a conversation was recorded as an indicator of engagement, or reversed, of avoidance.

Physiological measurement

Heart rate and skin conductance measurements were included to measure arousal elicited in the virtual world. The physiological measurements were done with a Mobi8 system from TMSi (see also Figure 4.4). Heart rate was recorded with an Xpod Oximeter, and the participants were requested to insert a finger into an adult articulated finger clip sensor. For skin conductance measurement two finger electrodes were used. An elevation in heart rate or skin conductance was regarded as an indicator for increased arousal.

Procedure

Prior to the experiment, participants were provided with an information sheet, and the procedure was explained to them. They were then asked to sign an informed consent form, and to fill in an information questionnaire and the PRCS questionnaire. Once immersed in the virtual environment, the participants were requested to have a conversation with the virtual lady. All the participants were exposed to all the six conditions, with six different topics in each condition. The topics were randomly assigned to the experimental conditions. The order of the conditions was counterbalanced across participants to control for possible systematic biases such as testing, learning, fatigue, or order effects between the conditions. The presence questionnaire, the DEQ and the SAM questionnaire were administered after each conversation with the virtual human. During the conversation, physiological data were recorded. The response of the participants was recorded with a web camera.

4.4 Results

The mean and standard deviation of the PRCS scores over all participants were $M = 9.12, SD = 4.15$. Taking the PRCS mean as a starting point, three groups of about equal size were created. However, as the PRCS index is a discrete score, it was not possible to create groups of exactly equal size. So, the division of participants over the groups we created was: the high confidence group (scores between 0 and 8, N=9), the medium confidence group (scores 9 or 10, N=7), and the low confidence group (scores between 11 and 16, N=8). Note that the medium size group covers only a relatively small PRCS range as a normal distribution centers around the mean. To reduce complexity, the reported analyses that include the PRCS groups as between-subjects variable, only include the two extreme groups, i.e., the low and high confidence group, and so, exclude the medium PRCS group[2]. The alpha level was set at .05 for all the tests.

As some of the depended variables deviated from normality, non-parametric analyses were conducted, including Mann-Whitney U tests for between-group comparisons, Wilcoxon Signed Ranks tests for paired comparisons, and linear-mixed-models analyses on aligned rank data for non-parametric factorial analyses (Wobbrock et al., 2011).

Six Mann-Whitney U tests (i.e., one per test condition) were conducted to compare the IPQ data from the 24 participants with the online IPQ dataset. The results suggested that a reasonable level of presence was obtained in the experiment as no significant difference was found between the overall median ($Mdn = 41, IQR = 14, n = 393$) of the IPQ online data set[3] for non-stereoscopic monitor and the median IPQ score in the `L-S-` ($Mdn = 41, IQR = 13, z = 0.53, p = .60$), `L0S0` ($Mdn = 45, IQR = 16.75, z = 1.87, p = .061$) and `LrSr` ($Mdn = 42.5, IQR = 16.25, z = 1.49, p = .14$) conditions. The measured level of presence was even significantly higher in the `L+S+` ($Mdn = 47, IQR = 12.5, z = 2.91, p = .004$), `L+S-` ($Mdn = 45.5, IQR = 18, z = 2.19, p = .028$) and `L-S+` ($Mdn = 44, IQR = 18.25, z = 2.30, p = .021$) conditions.

4.4.1 Positive versus negative synthetic emotion

To study the effect of the within-subjects factors regarding positive and negative synthetic emotions (hypothesis 1) in the listening and speaking phase, and the effect of the between-subjects factor regarding the low and high confidence group (hypothesis 3), several linear-mixed-models analyses on aligned

[2] In cases where conclusions with regarded to the hypothesis testing provided differ results, the results of the three level analyses are reported in the footnotes.

[3] The data was downloaded on April 3rd, 2013. http://www.igroup.org/pq/ipq/data.php

Table 4.2: Results of the Mixed-effect Model Analysis of Variance for Discussion Satisfaction

	Discussion Satisfaction
PRCS	$F(1,14) = 4.64, p = .049$
Listening	$F(1,49) = 3.47, p = .068$
Speaking	$F(1,46) = 33.69, p < .001$
PRCS × Listening	$F(1,48) = 0.78, p = .381$
PRCS × Speaking	$F(1,49) < 0.01, p = .981$
Listening × Speaking	$F(1,49) = 0.03, p = .862$
PRCS × Listening × Speaking	$F(1,49) = 0.03, p = .865$

rank data for non-parametric factorial analyses were conducted on participants' satisfaction and emotional state collected in the four conditions: `L+S+`, `L+S-`, `L-S+` and `L-S-`.

Dialog Satisfaction

The medians of the DEQ-satisfaction scores (with the IQR between brackets) for the `L+S+`, `L+S-`, `L-S+`, `L-S-` conditions were $1.67(1.33), 0.78(2.17), 1.33(1.78)$, and $0.67(2.06)$ respectively. The mixed-model analysis (see table 4.2) shows that the speaking behaviour of the virtual lady affected the participants' discussion satisfaction significantly; participants felt less satisfied with their conversation when the virtual lady showed negative emotions compared to positive emotions (which supports Hypothesis 1). The effect of the listening behaviour of the virtual lady on the discussion satisfaction approached a significant level. Similarly, participants seem less satisfied with the conversion when the virtual human showed negative instead of positive emotions during the listening phase (which tend to support Hypothesis 1). Less satisfaction was reported by participants with low speaking confidence ($Mdn = 0.28, IQR = 1.49$) compared to participants with high speaking confidence ($Mdn = 1.83, IQR = 1.89$), which supports Hypothesis 3.

Subjective Emotion

The SAM questionnaire was used to measure the participants' emotional state during their conversation with the virtual human. The medians and the interquartile ranges (in brackets) of the three emotional dimensions, i.e., valence, arousal and dominance for the `L+S+`, `L+S-`, `L-S+`, `L-S-` conditions are given in table 4.3. The results of the linear-mixed-model analysis on the aligned ranks data (see table 4.4) show that synthetic emotions in the speaking phase affected the participants' valence and dominance significantly. Participants reported a

Table 4.3: Median (IQR) of the SAM scores for high (High) and low confidence (Low) group.

	Valence			Arousal			Dominance		
Condition	Overall	Low	High	Overall	Low	High	Overall	Low	High
L+S+	6.0(2.0)	5.0(2.0)	7.0(2.0)	3.0(4.0)	3.0(3.0)	3.0(4.0)	5.0(5.0)	4.0(4.0)	6.0(3.0)
L+S-	4.0(3.0)	4.0(4.0)	5.0(4.0)	2.0(5.0)	2.5(4.0)	2.0(4.0)	4.0(3.0)	3.5(4.0)	4.0(5.0)
L-S+	6.0(3.0)	4.5(3.0)	6.0(2.0)	3.0(5.0)	3.5(4.0)	3.0(4.0)	5.0(4.0)	4.0(4.0)	6.0(4.0)
L-S-	4.0(3.0)	3.0(4.0)	5.0(3.0)	2.0(4.0)	3.5(3.0)	2.0(3.0)	4.0(3.0)	3.0(4.0)	4.0(3.0)

Table 4.4: Results of mixed-effect model analysis of variance for the sam scores.

	Valence	Arousal	Dominance
PRCS	$F(1,18) = 9.05, p = .008$	$F(1,17) = 1.45, p = .25$	$F(1,17) = 1.43, p = .25$
Listening	$F(1,45) = 1.50, p = .23$	$F(1,46) = 2.80, p = .10$	$F(1,50) = 0.04, p = .85$
Speaking	$F(1,47) = 21.0, p < .001$	$F(1,44) = 0.01, p = .91$	$F(1,50) = 5.69, p = .021$
PRCS × listening	$F(1,45) = 0.19, p = .67$	$F(1,47) = 8.06, p = .007$	$F(1,50) < 0.01, p = .98$
PRCS × speaking	$F(1,46) = 0.25, p = .62$	$F(1,45) = 2.34, p = .13$	$F(1,50) = 0.05, p = .83$
Listening × speaking	$F(1,45) = 0.38, p = .54$	$F(1,44) = 0.57, p = .46$	$F(1,50) = 0.23, p = .63$
PRCS × listening × speaking	$F(1,45) = 0.95, p = .33$	$F(1,44) = 0.41, p = .53$	$F(1,50) < 0.01, p = .95$

more positive emotional state and felt more dominant when the virtual human showed positive instead of negative speaking behaviour (which supports Hypothesis 1). On the contrary, the results did not show that the positive or negative emotions of the virtual lady during her listening phase affected the participants' emotional state. Furthermore, participants with low confidence ($Mdn = 3.50, IQR = 2.38$) reported lower valence scores than the participants with high confidence ($Mdn = 5.50, IQR = 1.75$) (which supports Hypothesis 3). The latter effect is visualized in Figure 4.5a. table 4.4 also shows a significant interaction between the PRCS groups and the listening behaviour of the virtual human on the reported arousal, which is visualized in Figure 4.5b. Especially, negative emotions expressed during the listening phase of the virtual human had a different impact on people with a low vs. high speaking confidence. Detailed analyses here only showed two trends: first, the low confidence participants tended to be more aroused ($z = 1.79, p = .074$) when the virtual human showed negative instead of positive listening behaviour, and second, low compared to high confidence participants reported more arousal ($z = 1.65, p = .099$) in the negative listening condition.

Dialog Length

The median (with the IQR between brackets) of the total talking time over all participants in seconds in the L+S+, L+S-, L-S+, L-S- conditions was

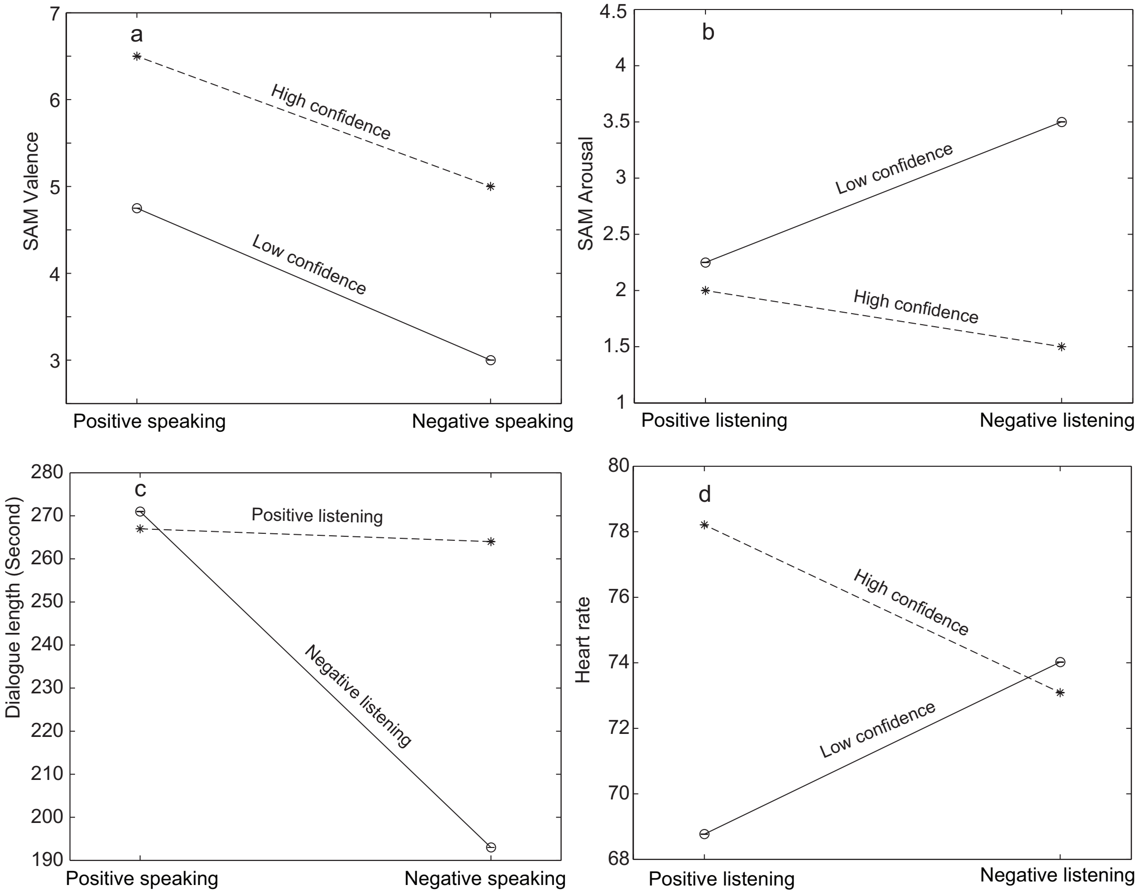


Figure 4.5: Median of SAM valence score (a), SAM arousal score (b), participants' dialog length (c), and heart rate (d).

$267.0(182.8), 264.0(214.0), 271.0(231.8)$, and $193.0(164.0)$ respectively (see also Figure 4.5c). Table 4.5 shows a significant main effect for the synthetic emotions expressed in the speaking phase. When the virtual human showed positive instead of negative speaking behaviour, the participants talked longer (which supports Hypothesis 1). Table 4.5 shows no significant main effect of the synthetic emotions in the listening phase on dialog length. In addition, Table 4.5 shows a significant interaction between the emotions expressed in the speaking and listening phase. As can be seen in Figure 4.5c, especially the combination of both negative speaking and listening behaviour resulted in a reduction of the speaking time, which was for example significantly ($z = 2.49, p = .013$) shorter than the speaking time in the positive listening and negative speaking condition.

Table 4.5: Results of mixed-effect models analysis of variance for the dialog length.

	Dialog length[4]
PRCS	$F(1, 18) = 0.19, p = .665$
Listening	$F(1, 45) = 2.82, p = .100$
Speaking	$F(1, 46) = 8.27, p = .006$
PRCS × listening	$F(1, 38) = 3.21, p = .081$
PRCS × speaking	$F(1, 40) = 0.05, p = .829$
Listening × speaking	$F(1, 42) = 5.14, p = .028$
PRCS × listening × speaking	$F(1, 41) = 0.02, p = .898$

Table 4.6: Results on the statistical analyses for the physiological measurements.

	Heart rate	Skin conductance
PRCS	$F(1, 17) = 2.51, p = .13$	$F(1, 15) = 0.57, p = .46$
Listening	$F(1, 49) = 1.44, p = .24$	$F(1, 39) = 4.59, p = .039$[5]
PRCS × listening	$F(1, 47) = 5.90, p = .019$	$F(1, 38) = 0.26, p = .61$
Speaking	$F(1, 49) = 0.18, p = .67$	$F(1, 40) = 0.23, p = .64$
PRCS × speaking	$F(1, 49) = 0.49, p = .49$	$F(1, 34) = 0.37, p = .55$
Listening × speaking	$F(1, 49) = 0.04, p = .85$	$F(1, 40) = 3.92, p = .055$
PRCS × listening × speaking	$F(1, 49) = 0.49, p = .49$	$F(1, 38) = 0.99, p = .33$

Physiological Measurements

The median (with the IQR between brackets) of the heart rate (averaged over the whole experimental time of one condition) in the `L+S+`, `L+S-`, `L-S+`, `L-S-` conditions was $73.72(16.30), 71.57(15.08), 73.24(18.81)$, and $72.93(14.48)$ respectively, while the median (in nano-Siemens and with the IQR between brackets) skin conductance (again averaged over the experimental time per condition) was $2526(2961), 2488(2401), 2585(2534)$, and $3794(2961)$ in the same conditions respectively. Table 4.6 shows a significant interaction between the PRCS groups and the listening behaviour on the heart rate data. As can be seen in Figure 4.5d, highly confident participants had a higher median heart rate than lowly confident participants when the virtual human expressed positive listening behaviour. This tendency approached the significance level ($z = 1.93, p = .054$). The corresponding detailed analysis also showed that the heart rate of only the low confidence group increased significantly ($z = 1.96, p = .050$) when the virtual lady changed her listening behaviour from positive to negative. Table 4.6 also shows a significant main effect for the listening behaviour on the participants' skin conductance, $F(1, 39) = 4.59, p = .039$. Participants sweated more when the virtual human expressed negative instead of positive listening behaviour.

[4] The interaction effect of listening and speaking was not significant ($F(1, 69) = 2.41, p = .13$) when the analysis was conducted using PRCS between-subjects variable with three levels.

Table 4.7: Median (with IQR between brackets) of the comparison between the speaking and listening phase, including the results of the corresponding Wilcoxon Signed Ranks tests ($n = 24$).

	$(S + L-) - (S - L+)$	Wilcoxon Signed Ranks tests
DEQ-discussion satisfaction	0.78(2.0)	$z = 2.86, p = .004$
SAM-valence	1.0(3.0)	$z = 2.97, p = .003$
SAM-arousal	0(1.0)	$z = 1.18, p = .24$
SAM-dominance	0(1.0)	$z = 1.48, p = .14$
Dialog length	21.0(125.0)	$z = 1.00, p = .32$
Heart rate	$-0.54(4.28)$	$z = 0.14, p = .89$
Skin conductance	9.7(70.20)	$z = 0.94, p = .35$

4.4.2 Listening vs. Speaking phase

To test whether synthetic emotions expressed in the speaking phase had more impact on the emotional valence and the satisfaction than emotions expressed in the listening phase (i.e., Hypothesis 2), the effects elicited in those two phases where contrasted against each other; in other words: speaking phase effect = listening phase effect. This contrast can be written as: [(`S+L-`) - (`S-L-`)] + [(`S+L+`) - (`S-L+`)] = [(`S-L+`) - (`S-L-`)] + [(`S+L+`) - (`S+L-`)], which is equivalent to (`S+L-`) - (`S-L+`) = `0`. Table 4.7 shows that the contrast value was significantly larger than zero for the score on discussion satisfaction and for the valence score, suggesting that the synthetic emotions had a larger impact during the speaking phase than during the listening phase (which supports Hypothesis 2).

4.4.3 Neutral vs. random

The median (with the IQR between brackets) of all dependent variables for all 24 participants in the neutral and random condition are shown in Table 4.8. The corresponding Wilcoxon Signed Rank tests show that participants were significantly less satisfied with their conversation when the virtual human showed random emotions ($Mdn = 1.0, IQR = 1.8$) instead of neutral emotions ($Mdn = 1.2, IQR = 1.4$), $z = 1.98, p = .048$ (which supports Hypothesis 4). Furthermore, participants felt themselves significantly less dominant (neutral: $Mdn = 5.0, IQR = 3.0$; random: $Mdn = 4.0, IQR = 4.0$) in the random condition, $z = 2.56, p = .011$.

[5] The effect of listening is not significant ($F(1, 59) = 2.96, p = .091$) when the analysis was conducted using PRCS between-subjects variable with three levels.

Table 4.8: Median (with IQR between brackets) of the scores for the random and neutral conditions, including the results of the corresponding Wilcoxon Signed Ranks tests.

	Neutral	Random	Wilcoxon Signed Ranks Tests
DEQ-discussion satisfaction	1.2(1.4)	1.0(1.8)	$z = 1.98, p = .048$
SAM-valence	5.0(3.0)	4.5(5.0)	$z = 1.87, p = .062$
SAM-arousal	2.0(4.0)	2.5(5.0)	$z = 0.94, p = .35$
SAM-dominance	5.0(3.0)	4.0(4.0)	$z = 2.56, p = .011$
Dialog length	298.5(163.3)	251.8(188.1)	$z = 1.03, p = .30$
Heart rate	70.9(13.7)	71.1(13.3)	$z = 0.14, p = .99$
Skin conductance	205.4(293.9)	203.7(270.5)	$z = 1.10, p = .27$

4.5 Discussion and conclusions

The analyses on the data for valence and discussion satisfaction suggest that positive compared to negative synthetic emotions expressed by a talking virtual human can elicit a more positive emotional state in a person, and can create more satisfaction towards the conversation. Therefore, we only found support for the first hypothesis in the speaking behaviour of the virtual human as no significant effect was found for the different emotions expressed by the listening virtual human. This dominance of the speaking phase over the listening phase was also hypothesised by the second hypothesis and confirmed by the data analyses since a larger effect on reported valence and discussion satisfaction was found for the synthetic emotions manipulated in the speaking phase compared to the listening phase of the virtual human. Besides the additional verbal channel to express emotions in the speaking phase, the participants might also have spent less attention to the virtual human when they were talking and the virtual human was listening. In human-human communication, the gaze of a listener is often fixed on the speaker, while the gaze of the speaker is only fixed on the listener when he or she begins or stops talking (Morris, 2002b).

Our findings also suggest that a conversation with a virtual human has clinical relevance as support was found for the third hypothesis. Participants with less speaking confidence obtained a more negative emotional state and were less satisfied with the discussion than participants with more speaking confidence. Although the experiment did not include individuals diagnosed with social anxiety disorder, social anxiety can be regarded as a continuous scale. Therefore these findings might generalise to the more extreme side of this scale. In this context, the results on the self-reported arousal and the dominance emotion dimensions, and on the physiological and behaviour measures are also interesting. For VRET to work effectively, it needs to be able to elicit fear. This emotion is a state of negative valence, high arousal, and low dominance. Negative speaking behaviour was not only able to create negative valence, but also to elicit a lower dominance level. This seems to replicate the findings reported by De Melo et al. (2012) on how people felt when negotiating with an angry virtual human. Additionally, the heart rate and subjective arousal of participants with low speaking confidence increased when they were confronted with negative instead of positive listening behaviour. As social anxiety is centred on the fear for negative social evaluation, these low confidence participants might have spent more attention to the virtual human when they were talking to see how it responded to them. We also observed more avoidance behaviour, i.e., reduced speaking time, when the virtual human expressed negative instead of positive speaking behaviour. This avoidance behaviour was even enhanced when negative speaking behaviour was combined with negative listening behaviour.

Our findings also show that a virtual human expressing randomly positive or negative emotions has a negative effect on the conversation satisfaction as compared to expressing neutral emotions. This result confirms the fourth hypothesis. In addition, the random behaviour made the participants feel less dominant. Again this seems to replicate reports on how negotiators felt when negotiating with someone that changed often from expressing anger to happiness (Sinaceur et al., 2013). These findings seem to have two practical implications. First, simply giving a virtual human the ability to express some random emotions may have a negative effect on the emotional state of the conversation partner. Second, if therapists in a simulated conversation environment change the emotions often it could reduce the conversation satisfaction.

Apart from the contributions, there are still a number of limitations to this study. First, although the study used a 3D virtual human with head and chest, full-body postures or gestures were not manipulated in this study. Considering that in recent decades more insights have become available on body expression (Gross et al., 2010; Kleinsmith and Bianchi-Berthouze, 2013), investigating the impact of full-body emotional expression of a virtual human is an interesting topic for future research, especially in relation to eliciting human emotions. Second, because of the language used by the virtual human, only Chinese participants were recruited, which might limit the generalisation of the findings to other nationalities. Still our conclusions seem to agree with findings of studies conducted with non-Chinese individuals. Third, only a sample of students from a technical university were recruited in this study, which also might limit the generalization of the findings to a larger more diverse population. Fourth, to have the human conversation partners perceive that they were negatively or positively evaluated by a virtual human, this study only used a limited set of facial expressions, i.e., basically expressing anger or happiness, where more negative and positive emotions exist. Future research could examine whether other negative emotions, such as sadness, fear, or frustration might also lead individuals to believe that they are negatively evaluated by a virtual human.

To conclude, the results of this paper show the effect of synthetic emotions in a conversation with a virtual human, especially when it is speaking. This suggests that designers who want to elicit emotions should especially focus on this phase of the conversation. The contributions of our study could help to improve the overall experience with simulated conversations, for example as part of a training, game, or psychotherapy.

Bibliography

Adler, M. J. (1997). *How to Speak How to Listen.* Touchstone; 1st Touchstone Ed edition.

Anderson, P. L., Jacobs, C., and 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., and 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.

Baylor, A. L., Ryu, J., and Shen, E. (2003). The effects of pedagogical agent voice and animation on learning, motivation and perceived persona. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pages 452–458.

Beidel, D. C., Turner, S. M., Jacob, R. G., and Cooley, M. R. (1989). Assessment of social phobia: Reliability of an impromptu speech task. *Journal of Anxiety Disorders*, 3(3):149–158.

Bradley, M. M. and Lang, P. J. (2007). The International Affective Digitized Sounds (IADS-2): Affective ratings of sounds and instruction manual. *Technical report B-3. University of Florida, Gainesville, Fl.*

Brinkman, W.-P., Hartanto, D., Kang, N., De Vliegher, D., Kampmann, I. L., Morina, N., Emmelkamp, P. M. G., and 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*, pages 1099–1102.

Brinkman, W.-P., Hattangadi, N., Meziane, Z., and Pul, P. (2011). Design and Evaluation of a Virtual Environment for the Treatment of Anger. In Richir, S. and Akihiko, S., editors, *Proceedings of Virtual Reality International Conference (VRIC 2011)*, pages 6–8, Laval, France.

Brinkman, W.-P., Van der Mast, C. A. P. G., and 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*, pages 83–95.

Brinkman, W.-P., Van der Mast, C. A. P. G., Sandino, G., Gunawan, L. T., and Emmelkamp, P. M. G. (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.

Broekens, J., Harbers, M., Brinkman, W.-P., Jonker, C., Van den Bosch, K., and Meyer, J.-J. (2012a). Virtual reality negotiation training increases negotiation knowledge and skill. In *IVA'12 Proceedings of the 12th international conference on Intelligent Virtual Agents*, pages 218–230.

Broekens, J., Qu, C., and Brinkman, W.-P. (2012b). Dynamic Facial Expression of Emotion Made Easy. In *Technical report. Interactive Intelligence, Delft University of Technology.* Technical report. Interactive Intelligence, Delft University of Technology.

Burleson, W. and Picard, R. W. (2007). Gender-specific approaches to developing emotionally intelligent learning companions. *Intelligent Systems*, 22(4):62–69.

Clark, D. A. and Beck, A. T. (2011). *Cognitive Therapy of Anxiety Disorders: Science and Practice.* The Guilford Press; 1st edition.

Core, M., Traum, D., Lane, H. C., Swartout, W. R., Marsella, S., Gratch, J., and Van Lent, M. (2006). Teaching negotiation skills through practice and reflection with virtual humans. *Simulation*, 82:685–701.

Cowell, A. J. and Stanney, K. M. (2003). Embodiment and Interaction Guidelines for Designing Credible, Trustworthy Embodied Conversational Agents. In Rist, T., Aylett, R., Ballin, D., and Rickel, J., editors, *4th International Workshop on Intelligent Virtual Agents IVA 2003*, volume 2792 of *Lecture Notes in Computer Science*, pages 301–309. Springer-Verlag.

Daly, J. A. (1978). The Assessment of Social-Communicative Anxiety Via Self-Reports: A Comparison of Measures. *Communication Monographs*, 45(3):204–218.

De Melo, C., Carnevale, P., and Gratch, J. (2012). The Effect of Virtual Agents' Emotion Displays and Appraisals on People's Decision Making in Negotiation. *Intelligent Virtual Agents*, pages 53–66.

Ekman, P. and Friesen, W. V. (1978). *Facial action coding system: A technique for the measurement of facial movement.* Consulting Psychologists Press, Stanford University, Palo Alto.

Emmelkamp, P. M. G. (2013). Behavior Therapy with Adults. In Lambert, M. J., editor, *Bergin and Garfield's Handbook of Psychotherapy and Behavior*, pages 343–392. John Wiley & Sons.

Gibb, J. R. (1961). Defensive Communication. *Journal of Communication*, 11(3):141–148.

Gratch, J. and Okhmatovskaia, A. (2006). Virtual rapport. In *Intelligent Virtual Agents*, pages 14–27.

Gratch, J., Rickel, J., Andre, E., Cassell, J., Petajan, E., and Badler, N. I. (2002). Creating interactive virtual humans: Some assembly required. *Intelligent Systems, IEEE*, 17(4):54–63.

Gregg, L. and Tarrier, N. (2007). Virtual reality in mental health: a review of the literature. *Social Psychiatry and Psychiatric Epidemiology*, 42(5):343–354.

Gross, M. M., Crane, E. A., and Fredrickson, B. L. (2010). Methodology for Assessing Bodily Expression of Emotion. *Journal of Nonverbal Behavior*, 34(4):223–248.

Gunawan, L. T., Van der Mast, C. A. P. G., Neerincx, M. A., Emmelkamp, P. M. G., and Krijn, M. (2004). Usability of therapist's user interface in virtual reality exposure therapy for fear of flying. In *In Proceedings of the Euromedia 2004*, pages 1–8.

Haring, M., Bee, N., and Andre, E. (2011). Creation and Evaluation of emotion expression with body movement, sound and eye color for humanoid robots. In *RO-MAN, 2011 IEEE*, pages 204–209. IEEE.

Irtel, H. (2007). PXLab: The Psychological Experiments Laboratory [online].

James, L. K., Lin, C.-Y., Steed, A., Swapp, D., and Slater, M. (2003). Social anxiety in virtual environments: results of a pilot study. *Cyberpsychology & Behavior*, 6(3):237–243.

Kleinsmith, A. and Bianchi-Berthouze, N. (2013). Affective Body Expression Perception and Recognition: A Survey. *IEEE Transactions on Affective Computing*, 4(1):15–33.

Klinger, E., Legeron, P., Roy, S., Chemin, I., Lauer, F., and Nugues, P. (2004). Virtual Reality Exposure in the Treatment of Social Phobia. *Studies in Health Technology and Informatics*, 99:91–119.

Konstantinidis, E. I., Hitoglou-Antoniadou, M., Luneski, A., Bamidis, P. D., and Nikolaidou, M. M. (2009). Using affective avatars and rich multimedia content for education of children with autism. *Proceedings of the 2nd International Conference on PErvsive Technologies Related to Assistive Environments - PETRA '09*, pages 1–6.

Krijn, M., Emmelkamp, P. M. G., Biemond, R., De Wilde de Ligny, C., Schuemie, M. J., Van der Mast, C. A. P. G., and De Ligny, C. D. (2004a). 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., and Biemond, R. (2004b). Virtual reality exposure therapy of anxiety disorders: a review. *Clinical Psychology Review*, 24(3):259–281.

Kulms, P., Kramer, N. C., Gratch, J., and Kang, S.-H. (2011). It's in Their Eyes: A Study on Female and Male Virtual Humans' Gaze. In *IVA'11 Proceedings of the 11th international conference on Intelligent virtual agents*, pages 80–92.

Kwon, J., Alan, C., and Czanner, S. (2009). A study of visual perception: social anxiety and virtual realism. In *Proceeding SCCG '09 Proceedings of the 25th Spring Conference on Computer Graphics*, pages 167–172.

Lang, P. J. (1995). The emotion probe. Studies of motivation and attention. *American Psychologist*, 50(5):372–385.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1999). International affective picture system (IAPS): Technical manual and affective ratings. Technical report, Gainesville University of Florida, Center for Research in Psychophysiology.

Lanyi, C. S., Stark, J. A., Kamson, M. E., and Geiszt, Z. (2011). Do we need high-scale flexibility in virtual therapies? *International Journal on Disability and Human Development*, 5(3):251–256.

Ling, Y., Brinkman, W.-P., Nefs, H. T., Qu, C., and 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.

Link, M., Armsby, P., Hubal, R. C., and Guinn, C. I. (2006). Accessibility and acceptance of responsive virtual human technology as a survey interviewer training tool. *Computers in Human Behavior*, 22(3):412–426.

Lok, B. (2006). Teaching communication skills with virtual humans. *Computer Graphics and Applications, IEEE*, 26(3).

Maldonado, H., Lee, J.-e. R., Brave, S., Nass, C., Nakajima, H., Yamada, R., Iwamura, K., and Morishima, Y. (2005). We Learn Better Together : Enhancing eLearning with Emotional Characters. In *Computer Supported Collaborative Learning 2005: The Next 10 Years!*, pages 408–417. Lawrence Erlbaum Associates, Mahwah, NJ.

Morris, D. (2002a). Contradictory Signals: Displaying two conflicting signals at the same time. In *Peoplewatching: The Desmond Morris Guide to Body Language*, pages 162–169. Vintage.

Morris, D. (2002b). Gaze Behaviour: Staring eyes and glancing eyes - the way we look at one another. In *Peoplewatching: The Desmond Morris Guide to Body Language*, pages 104–110. Vintage.

Morris, J. D. (1995). Observations : SAM The Self-Assessment Manikin An Efficient Cross-Cultural Measurement Of Emotional Response. *Journal of Advertising Research*, 35(6):63–68.

Opris, D., Pintea, S., Garcia-Palacios, A., Botella, C. M., Szamoskozi, S., and David, D. (2012). Virtual reality exposure therapy in anxiety disorders: a quantitative meta-analysis. *Depression and Anxiety*, 29(2):85–93.

Pan, X., Gillies, M., Barker, C., Clark, D. M., and Slater, M. (2012). Socially anxious and confident men interact with a forward virtual woman: an experimental study. *PLoS ONE*, 7(4):e32931.

Parsons, T. D. and 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.

Paul, G. L. (1966). *Insight Vs. Desensitization in Psychotherapy: An Experiment in Axiety Reduction*. Stanford University Press.

Pertaub, D.-P., Slater, M., and 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.

Phillips, G. C., Jones, G. E., Rieger, E. R., and Snell, J. B. (1997). Normative data for the personal report of confidence as a speaker. *Journal of Anxiety Disorders*, 11(2):215–220.

Powers, M. B. and Emmelkamp, P. M. G. (2008). Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders*, 22(3):561–569.

Powers, M. B., Francesca, N., Gresham, R., Jouriles, E. N., Emmelkamp, P. M. G., and Smits, J. A. J. (2013). Do conversations with virtual avatars increase feelings of social anxiety? *Journal of Anxiety Disorders*.

Qiu, L. and Benbasat, I. (2005). Online consumer trust and live help interfaces: The effects of text-to-speech voice and three-dimensional avatars. *International Journal of Human-Computer Interaction*, 19(1):37–41.

Reeves, B. and Nass, C. (1996). *The Media Equation.* Cambridge University Press.

Schelling, T. C. (1981). *The Strategy of Conflict.* Harvard University Press.

Schlosberg, H. (1941). A scale for the judgement of facial expressions. *Journal of Experimental Psychology*, 29:497–510.

Schroder, M. (2004). *Speech and Emotion Research: an overview of research frameworks and a dimensional approach to emotional speech synthesis.* PhD thesis, Saarland University.

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

Sinaceur, M., Adam, H., Van Kleef, G. A., and Galinsky, A. D. (2013). The advantages of being unpredictable: How emotional inconsistency extracts concessions in negotiation. *Journal of Experimental Social Psychology*, 49(3):498–508.

Slater, M., Pertaub, D.-P., Barker, C., and Clark, D. M. (2006). An experimental study on fear of public speaking using a virtual environment. *Cyberpsychology & Behavior*, 9(5):627–633.

Swartout, W. R. (2006). Simulators for human-oriented training. In *Proceedings - Winter Simulation Conference*, page 1202.

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

Ter Heijden, N. and 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., and 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*, pages 978–994.

Van Kleef, G. A. and De Dreu, C. K. W. (2010). Longer-term consequences of anger expression in negotiation: Retaliation or spillover? *Journal of Experimental Social Psychology*, 46(5):753–760.

Wandner, L., Hirsh, A., Torres, C., Lok, B., Scipio, C., Heft, M., and Robinson, M. (2013). Using virtual human technology to capture dentists' decision policies about pain. *Journal of Dental Research*, 92(4):301–305.

Wobbrock, J. O., Findlater, L., Gergle, D., and Higgins, J. J. (2011). The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only A NOVA Procedures. *CHI 2011 Session: Research Methods*, pages 143–146.

Wong, J. W.-E. and McGee, K. (2012). Frown More, Talk More: Effects of Facial Expressions in Establishing Conversational Rapport with Virtual Agents. In *IVA'12 Proceedings of the 12th international conference on Intelligent Virtual Agents*, pages 419–425.

Yee, N., Bailenson, J. N., and Rickertsen, K. (2007). A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1–10, New York, USA. ACM.

Chapter 5

The Virtual Bystanders

The effect of social evaluation, vicarious experience, cognitive consistency and praising:

students' beliefs, self-efficacy and anxiety to virtual bystanders in a language lesson

This chapter has been submitted to PloS one:
Chao Qu, Yun Ling, Ingrid Heynderickx, Willem-Paul Brinkman (submitted). Virtual bystanders in a language lesson: examining the effect of social evaluation, vicarious experience, cognitive consistency and praising on students' beliefs, self-efficacy and anxiety in a virtual reality environment. *PloS One.*

Bystanders in a real world's social setting have the ability to influence people's beliefs and behavior. This study examines whether this effect can be recreated in a virtual environment, by exposing people to virtual bystanders in a classroom setting. First participants (n = 26) witnessed virtual students answering questions from an English teacher, after which they were also asked to answer questions from the teacher as part of a simulated training for spoken English. During the experiment the attitudes of the other virtual students in the classroom was manipulated; they could whisper either positive or negative remarks to each other when a virtual student was talking or when a participant was talking. The results show that the expressed attitude of virtual bystanders towards the participants affected their self-efficacy, and their avoidance behavior. Furthermore, the experience of witnessing bystanders commenting negatively on the performance of other students raised the participants' heart rate when it was their turn to speak. Two-way interaction effects were also found on self-reported anxiety and self-efficacy. After witnessing bystanders' positive attitude towards peer students, participants' self-efficacy when answering questions received a boost when bystanders were also positive towards them, and a blow when bystanders reversed their attitude by being negative towards them. In addition, inconsistency, instead of consistency, between the bystanders' attitudes towards virtual peers and the participants was found to result in a larger change in the participants' beliefs about themselves. However, for beliefs participants held about others the findings were reversed. Finally the results also reveal that virtual flattering or destructive criticizing affected the participants' beliefs not only about the virtual bystanders, but also about the neutral teacher. Together these findings show that virtual bystanders in a classroom can affect people's beliefs, anxiety and behavior.

5.1 Introduction

Human's behavior, attitudes, emotions and cognition are extensively influenced by other people's opinions. People establish beliefs about these opinions regularly through one-to-one conversations with another person. Often, however, other individuals are present during such a social interaction; for example, fellow students in a class when a student talks to a teacher, colleagues in a meeting when someone talks to his or her boss, or people in a queue overhearing a person talking with someone at an information desk. Even though, these individuals, the so called bystanders, do not directly participate in the conversation, they may whisper or use nonverbal cues to express their opinion about what is being said.

Research showed that humans may be affected by the behavior of surrounding bystanders (Walster and Festinger, 1962) and that these bystanders may play an important role in human-human social interaction (Chaiken and Maheswaran, 1994). For example, behavior and judgments of a group of peers may influence an individual's cognition and judgment (Wetzel and Insko, 1982; Asch, 1951). This behavior may include words, intonations, gestures and facial expressions (Bailenson et al., 2005). Direct interaction with a virtual human (Villani et al., 2012; Qu et al., 2014, 2013) or virtual group (Anderson et al., 2013; Ling et al., 2012; Slater et al., 1999) has received considerable research attention, but this research has largely ignored the role of bystanders, in this case, virtual bystanders (an exception is the contribution of Lee (2011)). For applications that do want to offer the experience of social interaction in a virtual environment though, reports about bystanders in normal life suggest that virtual bystanders may have a relevant contribution in the experience of the social interaction. Take for example virtual reality exposure therapy (VRET) for the treatment of social anxiety disorder that is receiving increasing scientific and public attention (Villani et al., 2012; Anderson et al., 2013; Price et al., 2011). VRET is put forward as an alternative option for traditional exposure therapy in vivo because of its low cost, repeatability and convenient manipulation. In a recent controlled experiment, Anderson et al. (2013) found no difference in effectiveness between VRET and in vivo exposure therapy for treating social anxiety disorder. To be effective though, these virtual environments need to be engaging enough to activate anxiety in the patients (Foa and Kozak, 1986). The perception of negative human evaluation during social interaction is the main component to activate patients' social anxiety. The behavior and attitude of virtual bystanders can therefore play an important role and manipulating this may be a useful anxiety stimulus for therapists to control the intensity of patients' anxiety level.

The current study tries to address this gap in knowledge about virtual bystanders. An experiment was conducted to examine whether bystanders' judgments could influence a person's beliefs, self-efficacy and emotions during a virtual English lesson. The bystanders, i.e., virtual students, made either positive or negative comments, while other fellow virtual students or the human participant answered questions from a virtual teacher. The experiment was designed to address four hypotheses, of which the theoretical background is given in section 2 of this paper.

5.2 Theoretical Background

5.2.1 Bystander Evaluation

A bystander is a person who, although present at some event, does not take part in the event, and is often regarded as an observer or spectator. Although bystanders do not get involved in the event, their behavior may influence an individual's cognition and judgments. For example, Asch (1951) investigated the effect of majority opinions on individuals and found that people often modified their judgment in accordance with the majority. Also, the social facilitation theory claims that the presence of other people affects individual's performance, i.e., it enhances the individual's performance for well-practiced tasks, but impedes it for less familiar tasks (Geen, 1989). For example, Hunt and Hillery (Hunt and Hillery, 1973) showed that the presence of others reduced the number of errors produced by individuals learning an easy maze and increased it for those learning a difficult maze. Bystanders can also have an effect on each other. A well-known phenomenon studied in this context is the so-called bystander effect, referring to the observation that people are less likely to help a victim when other people are also present. The probability that a person actually provides help is inversely related to the number of other bystanders (Darley and Latane, 1968). Bystanders also play an important role in social comparison theory (Festinger, 1954), which argues that people evaluate their abilities and opinions by comparing it with others like them. This phenomenon occurs especially in situations where the evaluation is objectively unclear (Zitek and Hebl, 2007; Goldstein et al., 2008; Miller, 1984). For example, people are strongly influenced by the behavior of others when deciding whether to conserve energy in their homes (Schultz et al., 2007).

Likewise, people's perceived self-efficacy, i.e., the subjective probability that one is capable of executing a certain course of actions, has also been linked with verbal persuasion of others (Bandura, 1997). Evaluative feedback highlighting a person's capabilities raises efficacy (Schunk, 1984). Given the same level of performance, destructive criticism lowers perceived efficacy, whereas constructive criticism sustains or even boosts one's sense of perceived efficacy (Baron, 1988). Self-efficacy is a relevant concept for understanding people's behavior, since some studies have shown a strong relation between both (Bandura and Adams, 1977; Locke et al., 1984). As such, self-efficacy has also been related directly or indirectly to social anxiety. For example, Alden et al. (1992) found that people with low self-efficacy reported that they attended more to themselves and spent more time focusing on themselves during social interaction; hence, suggesting self-efficacy to be inversely related to self-focused attention. In addition, self-focused attention is one of the key symptoms of anxiety disorder and these symptoms have been reported to correlate with each other (Hope and Heimberg, 1985). Hope et al. (1987) found that socially anxious people were significantly more self-focused during social interaction than people who were not socially anxious. Other studies (Kashdan and Roberts, 2004; Thomasson and Psouni, 2010) have reported a direct inverse relation between self-efficacy and social anxiety.

Only recently has the idea of virtual bystanders received attention in the context of virtual environments. For example, Slater et al. (2013) tested the response of Arsenal supporters being bystanders to a violent argument in a virtual bar. They found that when the virtual victim was an Arsenal supporter instead of a person ambivalent towards the football club, the Arsenal supporters were more likely to physically and verbally intervene in the violent argument as they shared a common social identity with the virtual victim. Kozlov and Johansen (2010) were able to replicate in a virtual environment the inverse relation between the number of bystanders and the chance any person would intervene. They had people finding their way out of a virtual labyrinth that also included virtual characters that asked for their help. They found that people helped significantly less in situations with a large number of virtual bystanders compared to situations with no virtual bystanders. Also the social facilitation theory was studied in virtual reality. For example, Park and Catrambone (2007) found that for easy tasks people performed better in company with a virtual human than on their own, and they found the opposite effect for difficult tasks. Merely the presence of a virtual human in itself seems to cause this effect as this virtual human did not show any emotional expression and did not communicate with the participant during the task. Still, to the best of our knowledge, the effect of bystanders on individuals' dialog experience with a virtual human has not yet been studied empirically. Nevertheless, previous work that focused on direct interactions between a human and a virtual human has shown that a virtual audience (Pertaub et al., 2002) or a single virtual conversation partner (Qu et al., 2014) can effectively elicit higher or lower anxiety in a human speaker by expressing positive or negative emotions. Therefore, the current study investigates the effect of virtual bystanders expressing emotional behavior on an individual's experience by putting forward the first hypothesis: Positive compared to negative expressed attitudes by virtual bystanders towards a human speaker result in (H1a) higher self-perceived performance, (H1b) higher self-efficacy, and (H1c) less anxiety.

5.2.2 Modeling

The social cognitive theory (Bandura, 2001, 1977) suggests that people can learn from their observations, and use learned behavior when they are in the observed situation. Performances of observationally learned behavior are influenced by three major factors: personal standards of conduct, reward and punishment resulting from the observed, i.e., modeled behavior, and the similarity of the model (Bandura, 2001). People exhibit modeled behavior they find self-satisfying, but reject modeled behavior they personally disapprove (Bandura, 2001). People are motivated by the success of others who are similar to them. For example, the likelihood of learning increases when the models are of the same sex (Andsager et al., 2006), skill level (Meichenbaum, 1971) or have similar previous behaviors such as alcohol consumption (Andsager et al., 2006). People are more likely to perform the modeled behavior if it results in rewards instead of unrewarding or punishing effects. Bandura et al. (1963) found that children who observed an aggressive model being rewarded show more imitative aggression compared to children who observed a model being punished for the same aggressive behavior.

Modeling, also referred to as vicarious experience, has also been studied in virtual reality. Fox and Bailenson (2009), for example, let people observe a virtual lookalike or a dissimilar virtual person doing physical exercises. They found that either the reward of the virtual lookalike losing weight or the punishment of the virtual lookalike gaining weight was sufficient to encourage people to exercise significantly more than when observing these consequences affecting a virtual dissimilar person. However, what would happen if a person that is part of a group of bystanders, who are observing a conversation between two people, knows that he or she will be the next person having to have a conversation that will be observed by the same bystanders? As anticipation anxiety has been linked with performance anxiety (Brown and Stopa, 2007; Vassilopoulos, 2008), i.e., the fear to perform in front of others, this anticipated transition from a bystander to a person being observed might lead to anticipation anxiety especially if the individual witnesses negative consequences as a bystander for persons who are similar to him or her. Bandura (1997) suggests that when the vicarious experience includes positive consequences it may enhance self-efficacy. Hence, we expect that when bystanders witness positive feedback, their self-efficacy will raise and their anxiety will be reduced. Together this leads to the second hypothesis: Positive compared to negative expressed attitudes by virtual bystanders towards preceding virtual peer speakers results in (H2a) higher self-efficacy, and (H2b) less anxiety in a succeeding human speaker.

5.2.3 Consistency

Modeling can affect people's beliefs, but what happens to these beliefs if the real experience turns out to be inconsistent with the vicarious experience; for example, what happens if bystanders were positive towards peer student speakers, but later on negative towards the human speaker. In general, humans prefer consistency in behavior because of its perceptual simplicity (Heider, 1944). Consistency serves the need for coherence and effective action, and it is inherent to human nature as a result of neurophysiological processes and the capacity for logical reasoning (Ajzen, 2005). For example, in a study Somerville et al. (Somerville et al., 2006) let people perform a task while making social judgments and receiving fictitious feedback that was either consistent or inconsistent with their expectations. Their results demonstrated that participants had greater sensitivity to expectancy violations as compared to consistency in expectations, which suggests that people expected consistency in social exchange.

Inconsistency usually makes people psychologically uncomfortable (Festinger, 1957). In Festinger (1957) theory of cognitive dissonance, inconsistency between two beliefs exists when holding one belief conflicts with holding the other one. Inconsistency between cognitive elements such as beliefs and items of knowledge is assumed to enhance dissonance, which motivates the individual to change one or more cognitive elements to eliminate or reduce the magnitude of the dissonance. In other words, the theory of dissonance assumes a motivation for people to maintain consistency among their beliefs, feelings and actions. For example, when the individuals' actions conflict with their beliefs, they are expected to try to reduce the dissonance either by changing their beliefs or by changing their behaviors.

In a situation where bystanders first comment on the presentation of virtual peer speakers and later on a human speaker, inconsistency in these comments may force the human to change his or her belief much more extremely, than when the bystanders express exclusively positive or negative comments in both occasions. This leads to the third hypothesis (H3): Inconsistency in the bystanders expressed attitude towards virtual peer speakers and the human speaker leads to a larger change in belief than consistency in the bystander expressed attitudes.

5.2.4 Praise and destructive criticism

Up till now, the focus has only been on how bystanders affect beliefs people have about themselves. However, bystanders may also affect beliefs people have about the bystanders. For example, accumulated findings in the form of a meta-analysis (Gordon, 1996) support the claim that flattery has a positive influence on the people's judgment of the flatterer. The self-enhancement motive, i.e., people are motivated to evaluate themselves favorably and they respond positively by increased liking for people who flatter them, is suggested as a crucial factor underlying the positive effect of flattery (Gordon, 1996; Vonk, 2002). Colman and Olver (1978) showed that especially people with high self-esteem responded with a far greater liking for a flattering evaluator than for a more neutral evaluator that gave an assessments of their performance.

The effect of flattering was also studied in a virtual context. Reeves and Nass (1996) tested flattery by computers with a text-based user interface and found that individuals who were flattered by the computer performed better and liked the computer more than individuals who received no feedback or criticism from the computer. Fogg and Nass (1997) did a similar experiment and found that participants in the flattery condition reported more positive affect, better self-rated performance, more positive evaluations of the interaction and more positive perception of the computer, compared to the scores from participants in the generic condition. Johnson et al. (2004) extended the research of Fogg and Nass (1997) and also found that their participants reacted to flattery from a computer in a manner congruent with peoples' reactions to flattery from other humans, but only for participants with a high level of computer experience and not for participants with little computer experience. Consistently, Lee (2008) found that flattery led to more positive overall impressions and performance evaluations of the computer, but flattery also increased people's suspicion about the validity of the computer's feedback and lowered people's conformity to the computer's suggestions when they were answering the questions.

It seems therefore that bystanders' comments could also affect the beliefs people have about them. This therefore leads to the fourth and final hypothesis (H4): Beliefs about the bystanders correlate positively to the attitude bystanders expressed towards the human speaker.

5.3 Method

An experiment with a two-by-two within-subjects design existing of four conditions (as shown in table 5.1) was setup to test the four hypotheses. It included two within-subject factors: (1) the virtual bystanders' attitude towards the virtual student speakers who answered questions before the human speaker (i.e., the participant) got a turn to answer questions, and (2) the bystanders' attitude towards the human speaker. The bystanders' attitude could be either positive or negative, i.e., whispering either positive or negative remarks towards other bystanders and showing an angry or happy facial expression. Participants were exposed to all four conditions. To control for potential learning, order or fatiguing effects, the order of the four conditions was counterbalanced.

Table 5.1: Four experimental conditions with a different attitude of the virtual bystanders towards the virtual peer speakers and the participants

Condition	Bystanders' attitude towards virtual peer speakers (phase 1)	Bystanders' attitude towards the human (phase 2)
PP	Positive	Positive
NP	Negative	Positive
PN	Positive	Negative
NN	Negative	Negative

5.3.1 Participants

Twenty-six students (9 females and 17 males) from the Delft University of Technology participated in the experiment. Their age ranged from 20 to 30 years with the mean being 26.8 (SD = 2.5) years. All participants were non-native English speakers and they were all naive with respect to the hypotheses until they finished the experiment. Written informed consent was obtained from all participants prior to the experiment. Furthermore, for publication policy, the individual in this manuscript has also given written informed consent (as outlined in PLOS consent form) to publish case details. All participants received a small gift for their contribution. The experiment was approved by the Human Research Ethics Committee of the Delft University of Technology.

5.3.2 Measurements

The construct perceived performance, put forward in the hypotheses, was operationalized by considering the following indicators: (1) participants' rating of their own, virtual peers' and the teacher's performance, and (2) satisfaction with their own, virtual peers' and the teacher's performance. Anxiety was measured subjectively through the subjective units of discomfort (SUD) scale (Wolpe, 1969), physiologically, through skin conductance and heart rate, and behaviorally through speech length. In addition, the Personal Report Confidence as a Speaker (PRCS) questionnaire (Paul, 1966) and the Igroup presence questionnaire (IPQ) (Schubert et al., 2001) were used to measure the participants' general social anxiety and presence experienced in the virtual environment.

Personal Report of Confidence as a Speaker

The Personal Report of Confidence as a Speaker (PRCS) questionnaire (Paul, 1966) was used as a screening test for everyday experienced fear of speaking. It is a self-report questionnaire that assesses the behavioral and cognitive response to public speaking. It recorded whether participants agreed or disagreed (i.e., a binomial response) on 30 statements, for example "I dislike to using my body and voice expressively." The PRCS index was scored by counting the number of answers indicating anxiety. The PRCS index ranged from 0 to 30.

Presence questionnaires

Participants' sense of presence was also measured as recently a meta-analysis showed that anxiety experienced in a virtual environment is associated with presence (Ling et al., 2014). Participants were asked to complete the Igroup Presence Questionnaire (IPQ) (Schubert et al., 2001) at the end of the experiment to measure their experienced presence during the exposure in the virtual environment. IPQ consisted of 14 items rated on a seven-point Likert Scale. The scores on the 14 IPQ items were mapped onto three subscales, namely Involvement (i.e., the awareness devoted to the virtual environment), Spatial Presence (i.e., the relation between the virtual environment and the physical real world), and Realism (i.e., the sense of reality attributed to the virtual environment). The questionnaire also contained one item that assessed the general feeling of being in the virtual environment. The total score of IPQ was used in the data analysis to test whether the level of presence was sufficient to evoke an emotional response in the participants. The total score of IPQ ranged from 0 to 84.

Recently, Slater (2009) argued that presence at least has two independent components: place illusion and plausibility. Similar to physical presence, place illusion refers to the feeling of being in the virtual environment. Plausibility is the illusion that what is happening in the virtual world is really happening in spite of the knowledge that it is mediated technology. A high level of plausibility would elicit responses in the virtual environment similar to the ones in the real world. For VRET for social anxiety disorder, plausibility may be more relevant than place illusion. Therefore, for this experiment participants were asked to complete a created presence response scale (PRS), focusing on plausibility, using the following three items: (1) How often did you find yourself automatically behaving within the virtual English class as if it were a real English course? (2) To which extent was your overall behavior (what you said, emotional response and thoughts) like being in a real English course? (3) How much did you feel like being in a real English course?

Belief and experience questionnaire

The belief and experience questionnaire (BEQ) was specially made for this experiment and used to measure participant's beliefs about their own performance and that of the virtual peers and the teacher. The questionnaire also included questions with regards to satisfaction towards the performance of themselves, the virtual peers and the teacher, supportiveness of the virtual peers and the teacher, and self-efficacy. The formulation of self-efficacy question was based on self-efficacy question used in a study by Scherbaum (2006). All items were measured on scales ranging from 0 to 10.

The participant's experience of the lesson was measured on six semantic differential scales including unpleasant - pleasant, not relaxed - relaxed, aggressive - non-aggressive, uncomfortable - comfortable, impolite - polite and exhausting - energizing. All scales ranged from 0 to 10. The average score across the six scales was taken as an index for a participant's experience. The items of the BEQ questionnaire are shown in Appendix A.

Subjective units of discomfort

The 11-point scale of subjective unit of discomfort (SUD) was used to measure the perceived level of anxiety of the participants. A scores of 0 represented no fear and a score of 10 the highest level of fear an individual has ever felt in his or her life (Wolpe, 1969).

Physiological measurements

Physiological measurements, including heart rate and skin conductance, were included to measure elicited arousal during the virtual English lesson. The physiological measurements were done with a Mobi8 system from TMSi. To measure skin conductance two finger electrodes were used. Heart rate was recorded using an Xpod Oximeter, and the participants were requested to insert a finger into an adult articulated finger clip sensor. An elevation in heart rate or skin conductance was regarded as an indicator for increased arousal.

Speech length

Speaking time was suggested as a reliable behavioral measure to assess performance anxiety (Beidel et al., 1989). In an impromptu speech task, patients were asked to give a speech, and the length of the speech was taken as a reversed indicator of avoidance behavior. Anderson et al. (2013) also used the length of a participant's speech as a behavioral avoidance measure. Therefore, in this experiment the total time a participant talked during the discussion was recorded as an indicator of engagement, or reversed, of avoidance caused by anxiety.

Apparatus

As shown in figure 5.1, the virtual environment was displayed non-stereoscopically on a Sony HMZ-T1 Head-Mounted Display (HMD, 1280 × 720 pixels with 51.6° diagonal field of view) coupled to a three-degrees of freedom head tracker with a 500Hz update rate. Participants could freely look around to explore the virtual classroom, shown by means of a screenshot in figure 5.1b. Sound was played through embedded headphones. Besides the HMD, the participants wore a finger clip and two finger electrodes on their non-dominant hand for the Mobi8 system to record physiological data, including heart rate and skin conductance.

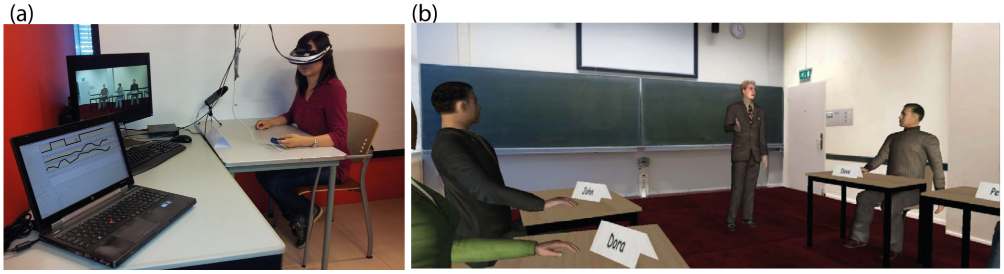


Figure 5.1: The experimental setup with (a) a participant doing the experiment, and (b) the participant's view of the virtual environment

The virtual environment was created using WorldViz's Vizard 3.0, and recreated an English lesson where a teacher asked students in turn general questions to practice their English conversation skill. Besides the participant, there were eight virtual fellow students sitting in the classroom: four males and four females. The classroom layout is shown in figure 5.2. The participant was always sitting on the third desk at the left side, while the position of the other virtual students was randomly assigned in each condition. The clothes and hair of the virtual students were always different in the four conditions to create the impression that each condition involved a different set of students.

In front of each virtual student was a name card, so the participant knew who was addressed by the teacher. The name of the participant was always 'Thomas' or 'Mary', depending on the gender of the participant. There was also a name card on the desk of the participant in real life to remind him or her of the

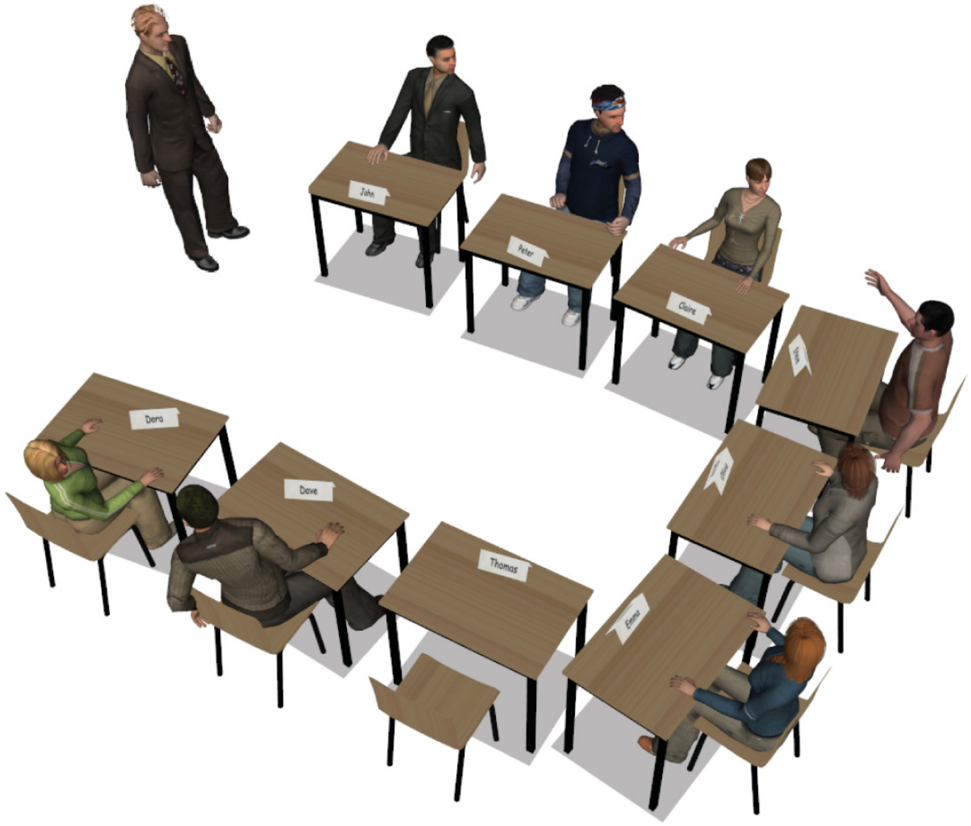


Figure 5.2: The layout of the virtual classroom where the participant was seated on the empty chair

temporary name in the virtual environment. The teacher was a well-dressed male around 40 to 50 years old. The voice actor of the teacher was a native English speaker, while the voice actors/actresses of the students were all non-native English speakers. A total of 28 open questions were recorded, seven for each condition. The teacher first posted four questions to different virtual students, randomly; when the last virtual student finished answering the forth question, the teacher asked the participant to answer that same question again, saying "Thomas/Mary, how about you?" and after that the last remaining three questions were all asked to the participant. The questions included examples as "What is the one thing that disgusts you and why?", "If you could go anywhere right now, where would that be and why?" and "What is the worst thing about being a grown-up and why?". So, these questions were formulated such that they had no clear objective evaluation criteria for the answers. All the eight voice actors/actresses of the virtual students were recorded while answering these questions spontaneously. So the teacher could ask anyone of the virtual students to answer his questions.

The virtual bystander students were able to show positive or negative behavior in the different experimental conditions. This behavior mainly consisted of facial expressions and whispering to each other, as illustrated in figure 5.3. Two facial expressions were used in this experiment: angry (see figure 5.3a) and happy (see figure 5.3b) to express negative or positive behavior respectively. The facial expression was achieved by a repeatable facial expression animation method, explained and evaluated in a previous study by Broekens et al. (2012). Different attitudes of the students were also expressed in their whispers. In the positive condition, virtual students whispered positively to each other; for example one student would say "Hey, this is a good answer!" and another bystander student would reply "Yes, a good one!", or the first student would say "I like it!", and another bystander would respond "I also like it!". In the negative condition, their whispers had a negative connotation; for example one bystander student could say "I don't like the answer!" and another student would replied "Me neither!", or a student would say "Boring!" and another student would reply "Yes, so boring!". So, all whispers focused on the content of the answers, and did not focus on the English formulation of the answer. A total of 36 pairs of positive or negative whisper dialogs were recorded for each of the eight virtual students, so participants could see whispering students on the left (figure 5.3c) or right (figure 5.3d) of them or in front of them (figure 5.3e). The whispers were designed to occur every 6-10 seconds after the virtual students or the participant started answering questions from the teacher. The questions towards the participants were triggered using speech detection. Three consecutive seconds of silence after the participant's answer triggered the teacher to give a neutral response such as 'ok' or 'all right', after which he would use a transition phrase such as 'next' or 'the next question' to introduce the next question. The speech detection was also designed to handle the situation that a participant would not say anything after the teacher posed a question. After 3 seconds of silence, the teacher would repeat the question and ask the participant to answer it again. In addition, to prevent a participant to give a very short answer to some of the questions, such as 'I don't know', the teacher would ask 'why' if the participant's answer was shorter than 5 seconds.

5.3.3 Procedure

Prior to the experiment, participants were provided with an information sheet, and the procedure was explained to them. They were then asked to sign an informed consent form, and to fill in a general information questionnaire, including SSQ and PRCS (see section 3.2 for more details). There were two phases in each condition: in the first phase, the virtual English teacher asked four virtual

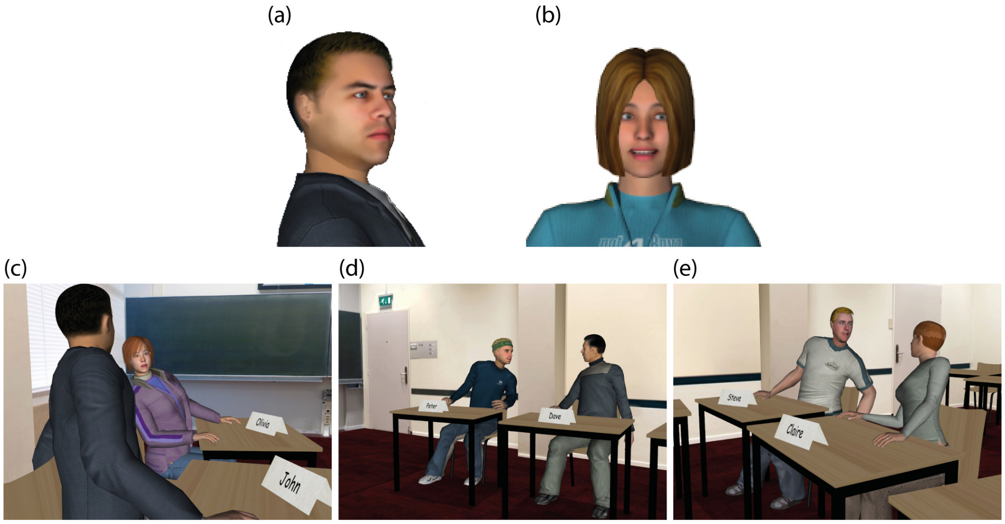


Figure 5.3: Facial expressions used in the experiment and virtual students (bystanders) whispering to each other (screenshots): (a) a virtual student showing an angry facial expression, (b) a virtual student showing a happy facial expression, (c) students whispering at the participant's left side, (d) students whispering in front of the participant, and (e) students whispering at the participant's right side.

peer students a question, and in the second phase, the teacher asked the participant four questions. The virtual bystanders' attitudes towards the virtual peer speakers and towards the participant were manipulated as either positive or negative according to table 5.1. At the start of each condition a pre-SUD score was obtained, whereas the BEQ, a post-SUD score and the PRS were administered at the end of each condition. After the participants experienced all the four conditions, the IPQ and SSQ were administered. Heart rate, skin conductance, and the length of a participant's answers were recorded during the experiment. The experimenter left the experimental room when a session started. Afterwards there was a debriefing session, in which the experimenter and the participant discussed the experiences and the experimenter explained to the participant the full details of the experiment. The whole experiment took about 50 minutes.

5.4 Results

The mean and standard deviation of the PRCS scores over all 26 participants were $M = 8.62, SD = 5.0$, indicating that the participants included in the experiment were generally socially confident. The IPQ data from the 26 participants suggested that a reasonable level of presence was obtained in the experiment as no significant difference was found between the IPQ total score of the online dataset[1] for non-stereoscopic HMD ($M = 45.73, SD = 7.98$) and the IPQ total score in the current experiment ($M = 46.35, SD = 9.86$) using an independent-samples t-test, $t(35) = 0.18, p = .86$. For each condition participants completed the BEQ, PRS, pre-SUD and post SUD questionnaires. This resulted in a set of 18 dependent variables expressing participants' beliefs about themselves (P1-P10), about the virtual other students (S1-S4) and about the teacher (T1-T4). The labels given here to the various dependent variables are consistently used in the various tables and in the remainder of the text of this paper. The data for the dependent variables were first normalized into z-scores for each participant across all items of a questionnaire and the four conditions. The following data analyses were based on the normalized scores.

5.4.1 Self-reported belief, experience, and anxiety

Mean and standard deviation of the 18 dependent variables for the four conditions are shown in table 5.2. The questionnaire measuring participants' experience in the virtual lesson had good reliability with Cronbach's α ranging from 0.71 to 0.88 across the four conditions.

The first step of the analysis consisted of eight repeated-measures ANOVAs using bystanders' attitudes towards (1) the virtual peer speakers and (2) towards the participants as two independent within-subject factors and as dependent variables: participants' rating of their own performance (P1), satisfaction with their own performance (P2), the virtual peers' satisfaction with their performance (P3), the teacher's satisfaction with their performance (P4), the beliefs whether virtual peers (P5) and the teacher (P6) liked them, the virtual lesson experience (P7) and self-efficacy (P8). The results are given in table 5.3, and show that the bystanders' positive attitude towards the participant compared to the conditions where the bystanders exhibited a negative attitude towards the participant, resulted in participants believing that peers and teacher were significantly more satisfied with their performance (P3 and P4) and liked them significantly more (P5 and P6), and resulted for the participants in a significantly more positive lesson experience (P7) and significantly more self-efficacy (P8). Although no significant effect for the bystanders' attitude on self-perceived performance (P1) and on the participants' satisfaction with their own performance (P2) was found, the p-value of 0.067 for the self-perceived performance (P1) approached the significant threshold of $\alpha = 0.05$. No significant main effect for bystanders' attitude towards their virtual peers was found for the eight dependent variables. Still a two-way interaction effect was found in the reported self-efficacy (P8), as illustrated in figure 5.4. When initially

[1] The data was downloaded on Oct 2nd, 2013. http://www.igroup.org/pq/ipq/data.php

Table 5.2: Mean and standard deviation of items of the BEQ and self-reported anxiety for the four experimental conditions

Measurements	PP	NP	PN	NN
The participants				
P1 Own performance	0.54 (0.66)	0.55 (0.63)	0.27 (0.78)	0.46 (0.64)
P2 Satisfaction with own performance	0.59 (0.63)	0.44 (0.72)	0.32 (0.83)	0.47 (0.76)
P3 Other students' satisfaction with your performance	0.85 (0.71)	0.98(0.69)	-0.90 (0.88)	-1.22 (0.73)
P4 Teacher's satisfaction with your performance	0.23 (0.58)	0.34 (0.71)	-0.15 (0.73)	-0.04 (0.56)
P5 Other students like you	0.95 (0.65)	0.98 (0.77)	-0.82 (0.83)	-1.11 (0.67)
P6 Teacher likes you	0.27 (0.57)	0.22 (0.74)	-0.03 (0.65)	-0.05 (0.63)
P7 Virtual lesson experience	0.59 (0.49)	0.42 (0.78)	-0.12 (0.60)	-0.28 (0.71)
P8 Self-efficacy	0.83 (0.47)	0.71 (0.73)	0.33 (0.81)	0.55 (0.62)
P9 SUD-post	-1.37 (0.94)	-1.03 (1.31)	-1.01 (1.15)	-1.03 (1.01)
P10 SUD post - SUD pre	-0.11 (0.80)	0.43 (0.95)	0.41 (0.59)	0.30 (0.74)
Other students				
S1 Other students' performance	0.46 (0.83)	0.42 (0.68)	-0.25 (0.83)	-0.17 (0.96)
S2 Participants' satisfaction with other students' performance	0.44 (0.57)	0.38 (0.51)	-0.28 (0.78)	-0.40 (0.88)
S3 Participants liking the other students	0.21 (0.70)	0.27 (0.62)	-0.51 (0.86)	-0.90 (0.85)
S4 How supportive were the other students towards you	1.00 (0.74)	1.27 (0.60)	-1.29 (0.69)	-1.53 (0.59)
The teacher				
T1 The teacher's performance	0.080 (0.62)	0.16 (0.64)	-0.10 (0.70)	-0.14 (0.67)
T2 Participants' satisfaction with teacher's performance	0.11 (0.65)	0.27 (0.46)	-0.09 (0.68)	- 0.28 (0.61)
T3 Participants liking of the teacher	0.14 (0.75)	0.05 (0.76)	-0.32 (0.82)	-0.33 (0.68)
T4 How supportive was the teacher towards you	-0.006 (0.82)	-0.09 (0.85)	-0.15 (0.84)	-0.50 (0.78)

the bystanders showed a positive attitude towards the virtual peer speakers, the participants' self-efficacy was significantly ($t(25) = 3.72, p = .001$) lower if the bystanders' attitude became negative instead of remaining positive when the participant was talking. However, when the bystanders initially showed a negative attitude towards the virtual peer speakers, no significant difference ($t(25) = 1.71, p = .099$) was found in participants' self-efficacy between conditions where the bystanders remained exhibiting a negative attitude or changed into exhibiting a positive attitude when the participant was talking.

Table 5.3 also shows the results of two repeated-measures ANOVAs with the same two within-subject factors on the self-reported anxiety at end of a session (P9, SUD post), and the change in self-reported anxiety (P10, SUD post - SUD pre). Although the analyses found no significant main effects, the effect of bystanders' attitude towards their virtual peers on the change in self-reported anxiety (P10) approached a significant level ($p = .063$). Related, the analyses revealed a significant interaction effect in the change in self-reported anxiety (P10), as also illustrated in figure 5.4. Participants reported significantly less change in anxiety in the condition where bystanders' attitude was positive towards both the virtual peers and the participant compared to all three other

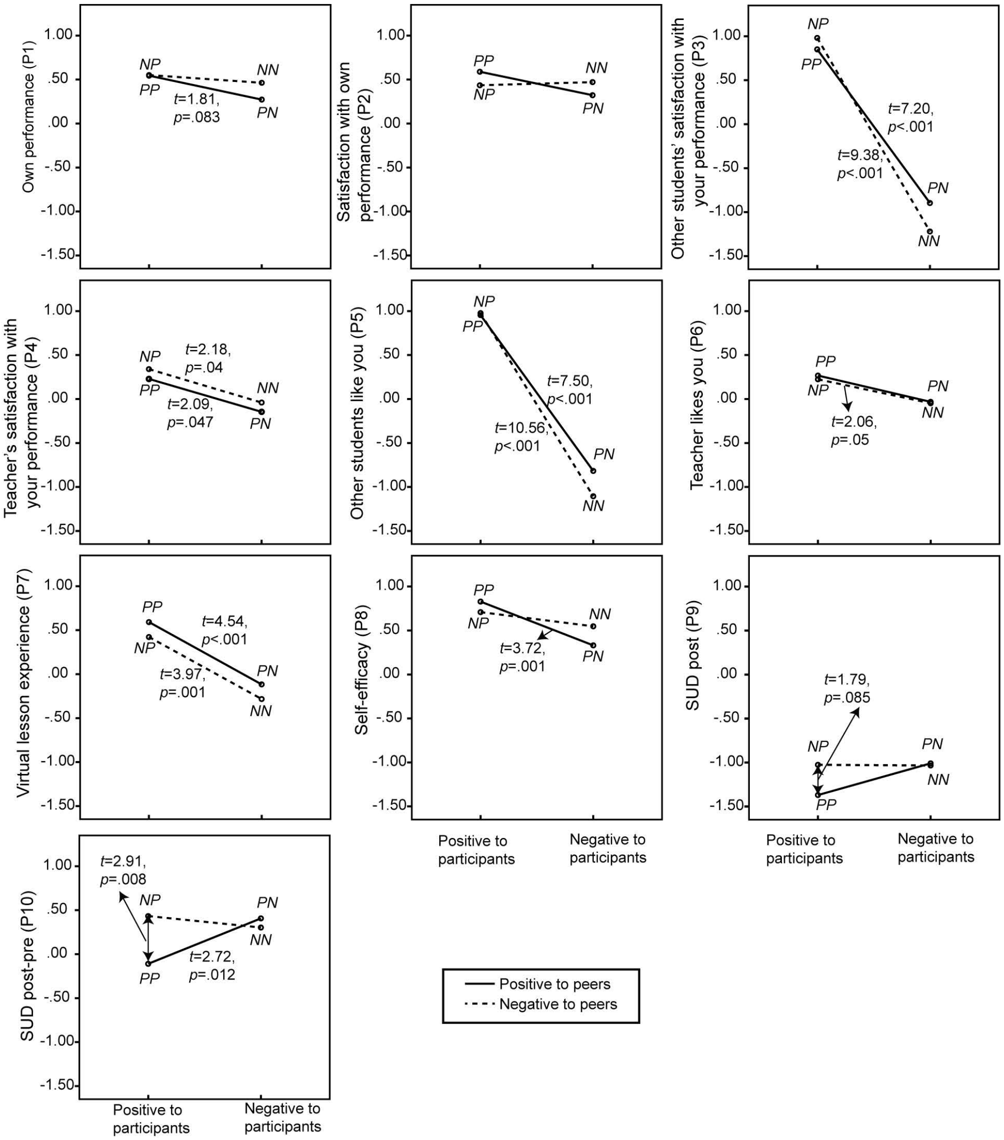


Figure 5.4: Results of the participants' self-related belief and experience questionnaire, and self-reported anxiety, including results of paired t-tests ($df = 25$).

conditions, i.e., (1) negative attitude towards peers and positive towards participant ($t(25) = 2.91, p = .008$), (2) positive towards peers and negative towards participant ($t(25) = 2.72, p = .012$), or negative towards both peers and participant ($t(25) = 3.01, p = .006$).

Besides focusing on their own, the BEQ also included questions about beliefs held towards bystanders and teachers. Eight repeated-measures ANOVAs were used, again with the same within-subject factors as independent variables,

Table 5.3: Results of the repeated-measures ANOVAs on items of the BEQ and self-reported anxiety

Measurements	Attitude towards					
	Participant		Peer speakers		Participant*Peer speakers	
	$F(1,25)$	p	$F(1,25)$	p	$F(1,25)$	p
The participants						
P1 Own performance	3.66	.067	1.07	.31	0.73	.40
P2 Satisfaction with own performance	0.81	.38	0.001	.98	2.70	.11
P3 Other students' satisfaction with your performance	95.25	$< .001$	0.97	.33	3.21	.085
P4 Teacher's satisfaction with your performance	12.03	.002	0.88	.36	$< .001$	.98
P5 Other students like you	97.53	.001	1.08	.31	2.60	.12
P6 Teacher likes you	6.85	.015	0.068	.80	0.014	.91
P7 Virtual lesson experience	23.45	$< .001$	2.86	.10	$< .001$	.98
P8 Self-efficacy	14.31	.001	0.20	.66	4.87	.037
P9 SUD post	1.29	.27	2.38	.14	1.53	.23
P10 SUD post- SUD pre	1.67	.21	3.79	.063	6.40	.018
Other students						
S1 Other students' performance	13.18	.001	0.034	.86	0.25	.62
S2 Participants' satisfaction with other students' performance	22.24	$< .001$	0.66	.42	0.125	.73
S3 Participants liking the other students	29.90	$< .001$	2.32	.14	3.97	.057
S4 How supportive were the other students towards you	269.56	$< .001$	0.028	.87	6.59	.017
The teacher						
T1 The teacher's performance	4.82	.038	0.11	.75	0.38	.55
T2 Participants' satisfaction with teacher's performance	7.76	.01	0.012	.91	4.11	.054
T3 Participants liking the teacher	11.86	.002	0.24	.63	0.18	.67
T4 How supportive was the teacher towards you	4.32	.048	3.21	.085	1.45	.24

and as dependent variables: the participants' rating of virtual peers' (S1) and teacher's performance (T1), participants' satisfaction with their performance (S2 and T2), how much the participant liked the peers (S3) and the teacher (T3) and their supportiveness (S4 and T4). The results of these analyses are also included in table 5.3 and further illustrated in figures 5.5 and 5.6. The results showed that bystanders' positive instead of negative attitude towards the participants resulted in significantly higher ratings for virtual peers' (S1) and teacher's (T1) performance, participant's satisfaction with these performances (S2 and T2), how much the participants liked them (S3 and T3) and their supportiveness towards the participant (S4 and T4). No significant main effect for bystanders' attitude towards the virtual peer speakers was found on any of these items. Still, the analysis did find a significant two-way interaction effect in the rating of peer students' supportiveness (S4). Figure 5.5 shows that participants believed that the virtual peers were more supportive towards them when peer students had been first negative to the student speakers and then positive towards them compared to the condition where the virtual peers had a positive attitude towards both the student speakers and participant,

$t(25) = 3.60, p = .001$. The opposite effect, however, was not found, i.e., a difference between conditions PN and NN, $t(25) = 1.27, p = .22$.

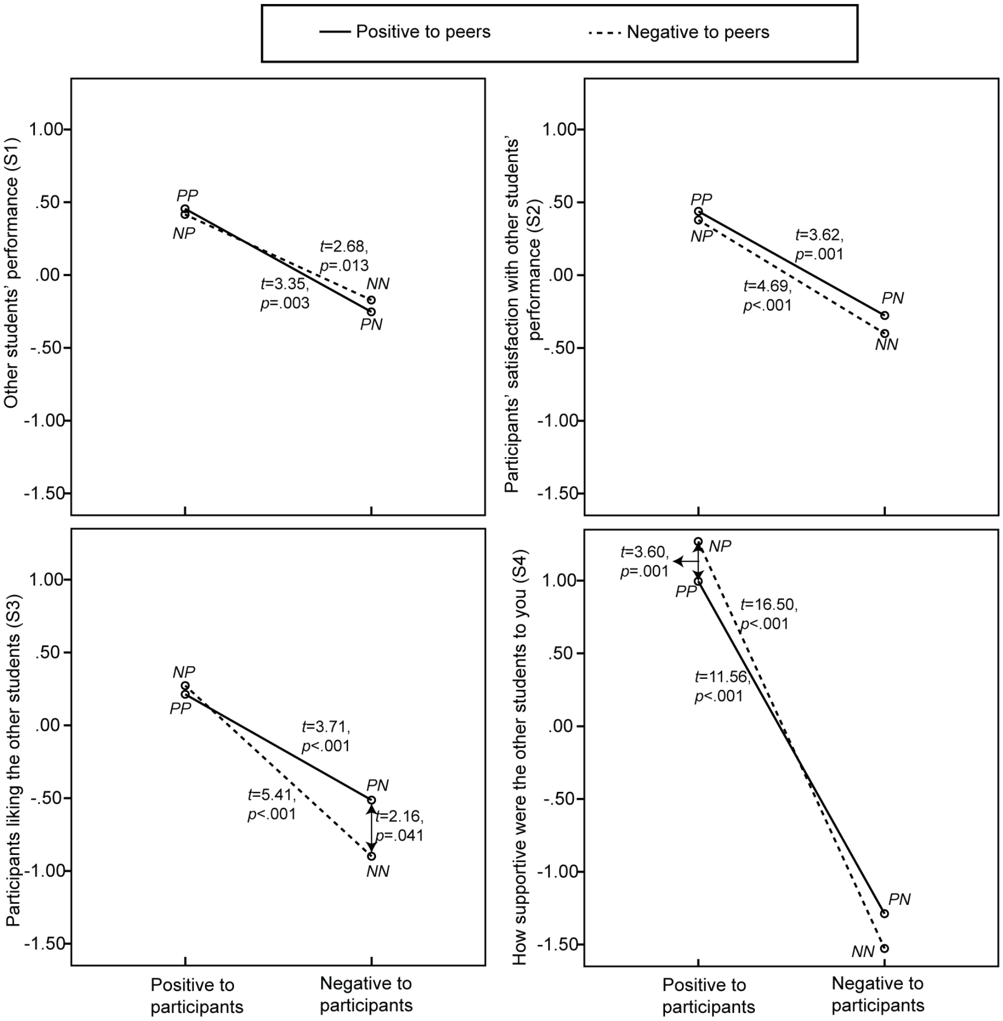


Figure 5.5: Participants' ratings of their beliefs regarding the virtual peers, including results of paired-samples t-tests ($df = 25$).

5.4.2 Presence response scale

The mean and standard deviation of scores on the presence response scale were 0.28 (0.69), 0.25 (0.58), -0.10 (0.76) and -0.15 (0.56) for PP, NP, PN and NN respectively. A repeated-measures ANOVA was conducted on participants' score on the presence response scale to test the effect of bystanders' attitude towards

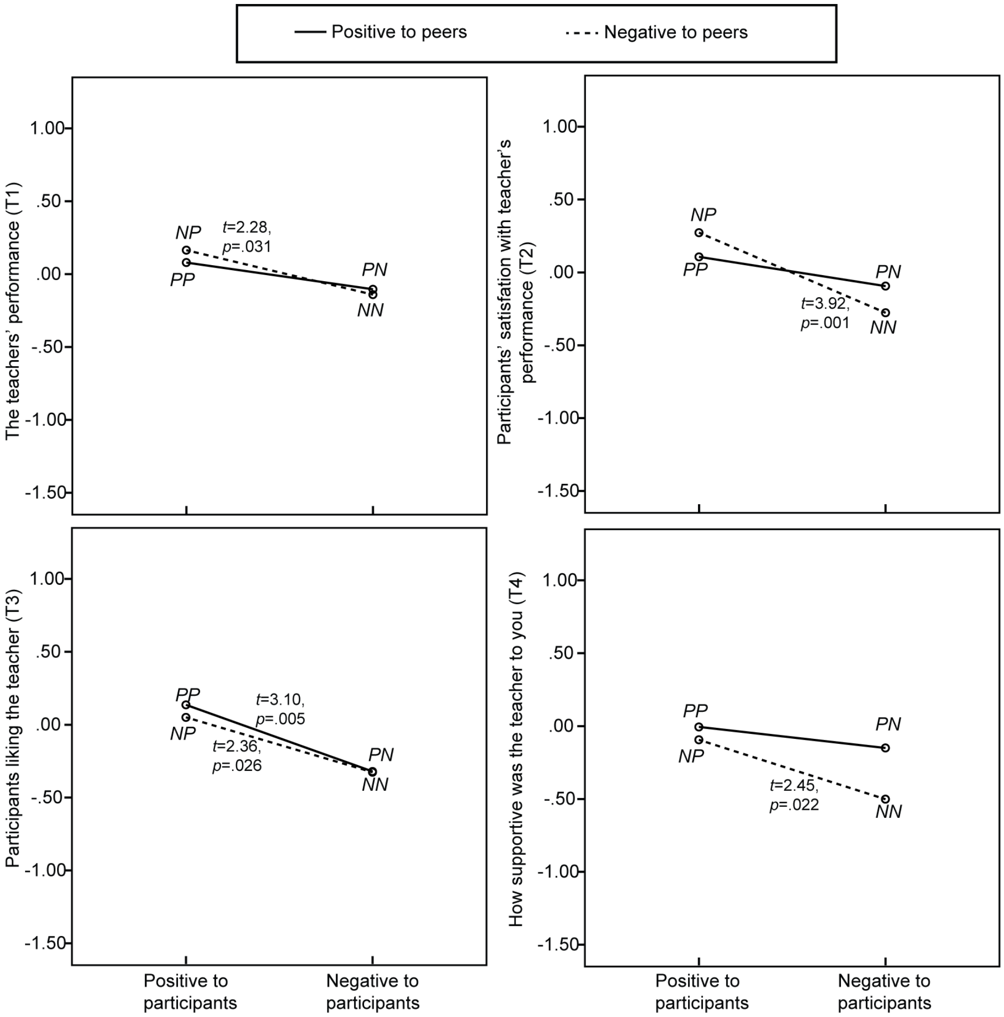


Figure 5.6: Participants' ratings of their beliefs regarding the teacher, including results of paired-samples t-tests ($df = 25$).

both virtual peer speakers and participants. The result showed a significant effect of bystanders' attitude towards the participants $F(1, 25) = 7.21, p = .013$ as participants rated their feeling of presence higher when bystanders had a positive instead of negative attitude towards them, see figure 5.7.

5.4.3 Speech length

The total speech length of participants' answers and the length of the answer to the first question were first normalized into z-scores for each participant across the four conditions.

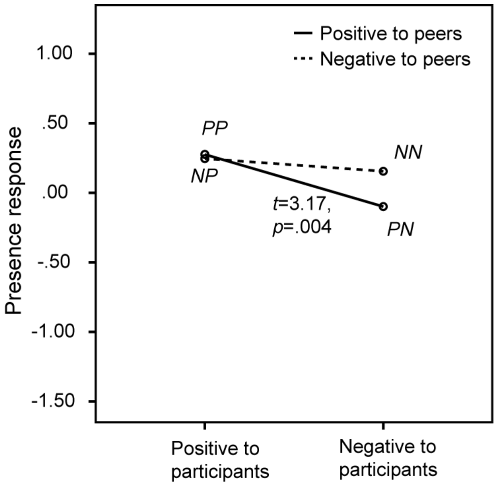


Figure 5.7: Participants' ratings of the presence response scale, including the result of a paired-samples t-test ($df = 25$).

Mean and standard deviation of participants' total speech length were 0.35 (0.63), 0.41 (0.67), -0.09 (1.00) and -0.67 (0.70) for PP, NP, PN and NN respectively, as also shown in figure 5.8. A repeated-measures ANOVA was conducted with the same two within-subject factors as before on participants' dialog length in each session. The results showed a significant main effect for the bystanders' attitude towards the participants, $F(1, 25) = 19.78, p < .001$. Participants gave longer answers when bystanders' attitude was positive instead of negative towards them. The main effect for the bystanders' attitude towards the virtual peer speakers approached a significant level, $F(1, 25) = 3.25, p = .084$. No significant two-way interaction was found, $F(1, 25) = 2.67, p = .12$.

The answer's length for the participants' first question was analyzed to examine the effect of the within-subject factors at the start of a participant's turn to speak. The means and standard deviations of the participants' speech length on the first question were 0.30 (0.74), 0.07 (0.90), -0.06 (0.97) and -0.31 (0.79) in the PP, NP, PN and NN conditions respectively. A repeated-measures ANOVA was conducted using the dialog length of the first question the participants answered as dependent variable. The result showed that the main effect for the bystanders' attitude towards the participants approached the significant level, $F(1, 25) = 3.23, p = .084$, with the answer's length being longer in the positive attitude condition than in negative one. Neither a significant main effect of the bystanders' attitude towards the virtual peer speakers ($F(1, 25) = 1.54, p = .27$), nor a significant two-way interaction ($F(1, 25) = 0.001, p = .93$) were found for the speech length on the first question.

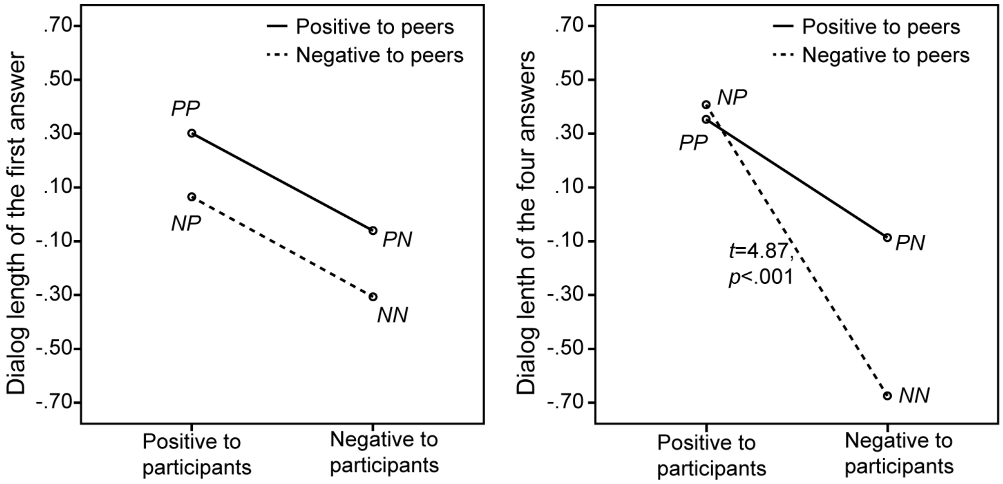


Figure 5.8: Participants' dialog length, including the result of a paired-samples t-test ($df = 25$).

5.4.4 Physiological Measurements

In contrast to the other data we collected, the physiological data also provided information when the participants were observing the virtual peer students answering questions. Therefore, we split our physiological data into two phases: the peer answering phase and the participant answering phase. The data on skin conductance were normalized into z-scores for each participant across eight moments, i.e., the two phases in each of the four conditions. As heart rate data was normally distributed, the analysis was conducted on the original data. One of the participant's physiological data was lost due to technical problems. Means and standard deviations of heart rate and skin conductance are shown in table 5.4.

Table 5.4: Means and standard deviations of heart rate and skin conductance across the two phases of the four conditions

	Phase 1 - Peers answering				Phase 2 - Participants answering			
	PP	NP	PN	NN	PP	NP	PN	NN
Heart rate	71.83	71.35	72.03	71.76	72.68	73.18	72.78	73.76
	(10.59)	(9.99)	(9.31)	(11.93)	(10.34)	(10.41)	(10.08)	(12.05)
Skin conductance	-0.29	0.12	0.11	-0.12	-0.16	0.18	0.17	-0.017
	(1.07)	(0.82)	(0.88)	(0.81)	(0.98)	(1.05)	(0.82)	(1.04)

A repeated-measures ANOVA was conducted using heart rate as dependent variable, and the phases of the lesson, the bystanders' attitude towards virtual peer speakers and participants as independent variables. The results, given in table 5.5, showed a significant effect for the phase on participants' heart rate with an increase in participants' heart rate in the second phase, where they answered questions. The analysis also found a two-way interaction effect between phase and bystanders' attitude towards the virtual peer speakers, as also illustrated in figure 5.9a. A significant increase in heart was found between the two phases only when in the first phase the participants observed a negative instead of positive attitude towards the virtual peers, $t(24) = 3.10, p = .005$. If they first observed a positive attitude towards the virtual peers no significant difference was found between the two phases, $t(24) = 1.54, p = .14$.

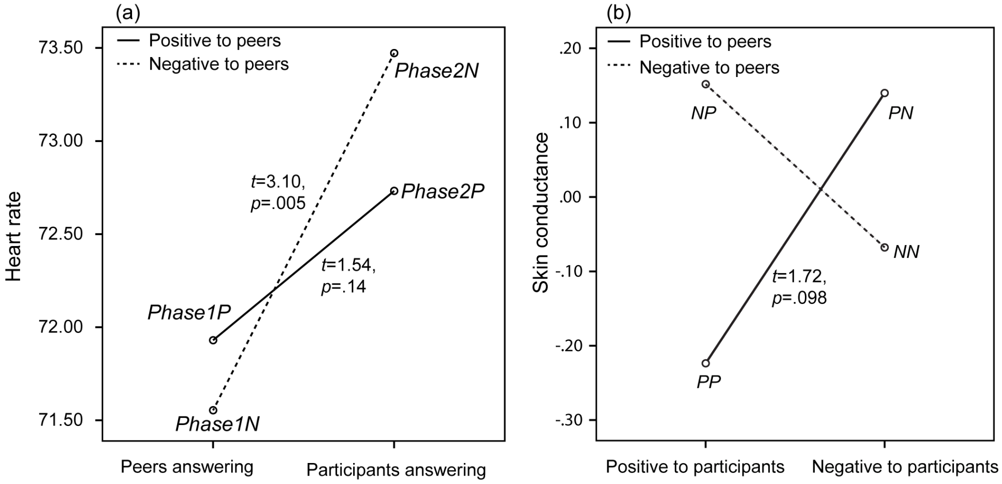


Figure 5.9: Participants' heart rate when peers or participants were answering questions, including results of paired-samples t-tests ($df = 24$).

A similar repeated-measures ANOVA was conducted with skin conductance as dependent variable. Although the results did not show any significant effect, Table 5.5 shows a two-way interaction effect between bystanders' attitude towards the participants and virtual peer speakers approaching significance ($p = .08$). As figure 5.9b shows when bystanders first expressed a positive attitude towards the virtual peer speakers, participants sweat more when after this bystanders expressed a negative attitude instead of a positive attitude towards them. Note again that this difference was only approaching an significant level ($t(24) = 1.72, p = .098$).

5.4.5 Consistency

The following contrast was examined $(PP - NN)^2 = (NP - PN)^2$ for all the belief data collected to test the third hypothesis that inconsistency compared to

Table 5.5: The results of repeated-measures ANOVAs for heart rate and skin conductance

Independent variables	Heart rate		Skin conductance	
	$F(1,24)$	p	$F(1,24)$	p
Phase	6.675	.016	0.249	.62
Attitude to participants	0.282	.60	0.195	.66
Attitude to peers	0.132	.72	0.266	.66
Phase * participants	0.003	.96	0.004	.95
Phase * peers	6.370	.019	0.003	.96
Participants * peers	0.058	.81	3.35	.08
Phase * participants * peers	0.055	.82	0.044	.84

consistency in the bystanders' expressed attitude towards virtual peer speakers and the human speaker lead to a larger change in belief. The results of paired-samples t-tests, given in table 5.6, showed that participants changed their belief about their own performance ($t(25) = -2.06, p = .05$) and the teacher's satisfaction with their own performance ($t(25) = -2.18, p = .04$) more extremely when the virtual bystanders showed inconsistent instead of consistent attitudes towards the virtual peers and the participant. Interestingly, a significant larger difference was found between the consistent conditions, i.e., $(PP-NN)^2$, compared to the inconsistent conditions, i.e., $(NP-PN)^2$, for the participants' beliefs about the performance of the virtual peers ($t(25) = 2.07, p < .05$) and the participants' satisfaction with the other students' performance ($t(25) = 2.78, p = .01$).

5.5 Discussion and conclusions

Given these results a number of conclusions can be drawn. First, virtual bystanders exhibiting positive instead of negative attitude towards the participants, make the participants to hold more positive beliefs about their own self-efficacy (supports H1b) and to behave more engaging by giving longer answers, i.e., showing less avoidance behavior which is interpreted as a manifestation of less anxiety (support H1c), confirming part of the hypothesis about the influence of bystanders' attitude. Also, the two-way interaction effect on the self-reported anxiety showing that bystanders' consistent positive attitude towards both the peer speakers and the participants evoked the lowest level of anxiety in the participants, supports H1c. Although no significant effect for bystanders' attitude toward the participants on participants' perceived performance (H1a) was found, the effect approached significance in the hypothesized direction.

Second, as predicted the participants seem to have experienced anticipation

Table 5.6: Means and standard deviations of consistent $(PP-NN)^2$ and inconsistent $(NP-PN)^2$ conditions, including the results of paired-samples t-tests

Measurements	$(PP-NN)^2$	$(NP-PN)^2$	Paired-samples t-tests	
	Mean (SD)	Mean (SD)	t	p
The participants				
P1 Own performance	0.26(0.32)	0.72(1.27)	t(25) -2.06	.05
P2 Satisfaction with own performance	0.39(0.93)	0.97(1.25)	t(25) -2.00	.057
P3 Other students' satisfaction with your performance	5.43(3.85)	4.90(3.78)	t(25) 0.70	.49
P4 Teacher's satisfaction with your performance	0.43(0.59)	1.14(1.75)	t(25) -2.18	.04
P5 Other students like you	5.22(4.13)	4.95(4.51)	t(25) 0.28	.78
P6 Teacher likes you	0.66(0.91)	0.73(1.06)	t(25) 0.29	.77
P7 Virtual lesson experience	1.44 (2.10)	1.17 (1.74)	t(25) 0.60	.55
P8 Self-efficacy	0.40(0.80)	0.80(1.68)	t(25) -1.15	.26
Other students				
S1 Other students' performance	2.06(2.67)	1.03(1.22)	t(25) 2.07	.049
S2 Participants' satisfaction with other students' performance	2.01(2.03)	1.02(1.19)	t(25) 2.78	.01
S3 Participants like the other students	2.30(2.30)	1.63(2.08)	t(25) 1.38	.18
S4 How supportive were the other students to you	7.29(4.79)	7.34(4.44)	t(25) -0.05	.96
The teacher				
T1 The teacher's performance	0.51(0.70)	0.51(0.87)	t(25) 0.01	.99
T2 Participants' satisfaction with teacher's performance	0.82(1.35)	0.69(1.09)	t(25) 0.41	.68
T3 Participants Like the teacher	0.76(1.41)	0.76(1.29)	t(25) 0.01	.99
T4 How supportive was the teacher to you	0.96(1.72)	0.92(1.44)	t(25) 0.09	.93

anxiety as their heart rate increased when they had to actively answer the teacher's question after passively observing bystanders exhibiting especially a negative attitude towards the virtual peer speakers (supports H2b). The effect of bystanders' attitude towards the peer speakers was also approaching significant on participants' rating of anxiety (so, also supports H2b). More support for the second hypothesis was found in the two-way interaction effects on the self-reported anxiety (H2b) and self-efficacy (H2a). For self-reported anxiety, it seems that any of the two types of negative attitude displayed increased anxiety. Still being negative at the same time towards both the virtual peer speakers and the participants did not seem to elicit more anxiety than the conditions where bystanders showed negative attitudes only towards the virtual peer speakers or the participants. For self-efficacy, it seems that once the participants had witnessed the bystanders' positive attitude to their virtual predecessors, a positive attitude towards them gave their self-efficacy a boost while a negative attitude a blow.

A third conclusion that can be drawn relates to the effect of consistency and inconsistency between bystanders' attitudes. The third hypothesis predicted that inconsistency causes larger changes in participants' beliefs than consistency. This was observed for the participants' beliefs about their own performance and the teacher's satisfaction with their performance. However, when it came to the beliefs the participants held about the performance of other students and the participants' satisfaction with other students' performance, the results were opposite to the prediction. Therefore the third hypothesis should be rejected or at least needs to be limited to only people's beliefs regarding their self-image. The finding about consistency for people's beliefs regarding other people's ability may be explained by a combination of flattering, criticism and also modeling. So, when students were flattering the other students, the participants were more inclined to believe this when the students flattered them as well. Similarly, when students were destructively criticizing the other students, the participants were more inclined to believe this when they also were destructively criticized.

Support for the fourth hypothesis about virtual flattering and destructive criticism was also found. When other students were flattering instead of criticizing destructively the participants, the participants rated the students' performance higher, were more satisfied with their performance, liked them more, and found them more supportive. Interestingly, this effect also rubbed off to the neutral teacher as similar effects were also found for participants' beliefs about the teacher. Still, instead of simply rubbing off, in the debriefing some participants mentioned that they regarded the neutral stance of the teacher as inappropriate, since they expected him to intervene when students were openly making negative comments.

Besides the predicted effects, the experiment also revealed some unexpected findings when it came to participants' feelings of presence as participants rated their presence higher when the bystanders exhibited a positive instead of a negative attitude towards them. This could again be a case of a rubbing off effect, i.e., towards the quality of the virtual reality environment. The cognitive dissonance theory (Festinger, 1957), however, also offers an explanation. As participants experienced inconsistency between the bystanders' negative attitude towards them and the positive self-image participants probably held, participants could have resolved this inconsistency by changing their belief about the credibility of the virtual experience; in other words, they could start questioning the plausibility, regarding it as dissimilar to their beliefs of the real world (Slater, 2009; Pan et al., 2012).

Like any empirical study, this experiment also had a number of limitations that should be noted. First, the task in the virtual environment was quite familiar to the participants which might have limited the effect of vicarious experience as vicarious experiences are a particularly valuable source of reassurance mainly when people are unsure about their own capabilities (Bandura, 1997; Takata and Takata, 1976). Future work could therefore test the vicarious experience in scenarios where individuals lack direct knowledge of their own capabilities, such as in a virtual teaching lesson or acting lesson, where they would rely more heavily on modeled indicators. Second, the participants were asked to answer questions more often than the other virtual students, which could have lowered their vicarious experience, but gave them more exposure to the bystanders' direct evaluation. It would, however, also be interesting to study the effect of virtual bystanders in a more normal full-length English lesson of around 45 minutes where people spend more time observing others instead of speaking themselves. Third, to avoid interrupting the flow of the experience no self-reported data were collected directly after the participants witnessed their virtual students' answers. However, these data would give insight into the effect of vicarious experience on self-efficacy and anticipation anxiety just before it was the participants' turn to speak. Still, the collected heart rate data did provide some insight into their anxiety. Fourth, neither the attitude of the bystanders nor the response of the teacher changed as a reaction towards the participants' performance. Still other have shown that providing positive or negative feedback in a dialog can affect people's emotion and behavior (Qu et al., 2014; Hartanto et al., 2014). Therefore, making bystanders or the teacher adapt their attitude to the performance of participants might affect participants' motivation as they could experience that their effort could have an impact on their environment. Finally, this experiment only recruited university students as participants, which makes it a small university sample study. It would be interesting to study how virtual bystanders would affect other groups of people, such as patients that suffer from social anxiety disorder.

The main contribution of the research presented is to establish insight into the effect of virtual bystanders in a virtual reality environment. The attitude expressed by them can have a clear effect on people's beliefs, self-efficacy, and anxiety. Therefore, manipulating the virtual bystanders' attitude could give therapists a tool to control the exposure in virtual reality environments for the treatment of social anxiety disorder. Another contribution of this work is the insight it provides into classroom dynamics. The simulation in the virtual classroom suggests that fellow classmates exhibiting a positive attitude towards each other leads students to act more engaging, to have more self-efficacy and to experience less anxiety, while a negative attitude could have a detrimental effect on all these aspects. Although teachers might take a neutral stance towards the class attitude, it still forms students' beliefs towards them. To conclude, the virtual bystander seems to have a clear ability to have an impact on the social experience in virtual environments that seem to correspond to what people experience in everyday life.

5.6 Appendix A: The belief and experience questionnaire (BEQ)

With regard to the English lesson in the last session:
P1: How would you rate your performance? 11-point scale from 0 (very bad) to 10 (very good)
P2: How satisfied are you with your performance? 11-point scale from 0 (highly unsatisfied) to 10 (highly satisfied)
P3: How satisfied do you think the other students were with your performance? 11-point scale from 0 (highly unsatisfied) to 10 (highly satisfied)
P4: How satisfied do you think the teacher was with your performance? 11-point scale from 0 (highly unsatisfied) to 10 (highly satisfied)
P5: How much do you think the other students like you? 11-point scale from 0 (not at all) to 10 (very much)
P6: How much do you think the teacher likes you? 11-point scale from 0 (not at all) to 10 (very much)
P8: After the last session, how confident and competent do you feel now in giving answers in an English lesson in real life? 11-point scale from 0 (not confident at all) to 10 (very confident)

S1: How would you rate the performance of the other students? 11-point scale from 0 (very bad) to 10 (very good)
S2: How satisfied are you with the performance of the other students? 11-point scale from 0 (highly unsatisfied) to 10 (highly satisfied)
S3: How much do you like the other students? 11-point scale from 0 (not at all) to 10 (very much)
S4: How supportive were the other students when you were giving answers? 11-point scale from 0 (very unsupportive) to 10 (very supportive)

T1: How would you rate the performance of the teacher? 11-point scale from 0 (very bad) to 10 (very good)
T2: How satisfied are you with the performance of the teacher? 11-point scale from 0 (highly unsatisfied) to 10 (highly satisfied)
T3: How much do you like the teacher? 11-point scale from 0 (not at all) to 10 (very much)
T4: How supportive was the teacher when you were giving answers? 11-point scale from 0 (very unsupportive) to 10 (very supportive)

P7: Please indicate how you experienced the English lesson in the last session. I experienced the lesson as:

P7.1: 11-point scale from 0 (unpleasant) to 10 (pleasant)
P7.2: 11-point scale from 0 (not relaxed) to 10 (relaxed)
P7.3: 11-point scale from 0 (aggressive) to 10 (non-aggressive)
P7.4: 11-point scale from 0 (uncomfortable) to 10 (comfortable)
P7.5: 11-point scale from 0 (impolite) to 10 (polite)
P7.6: 11-point scale from 0 (exhausting) to 10 (energizing)

Bibliography

Ajzen, I. (2005). *Attitudes, Personality, and Behavior.* Open University Press.

Alden, L. E., Teschuk, M., and Tee, K. (1992). Public self-awareness and withdrawal from social interactions. *Cognitive Therapy and Research*, 16(3):249–267.

Anderson, P. L., Price, M., Edwards, S. M., Obasaju, M. a., Schmertz, S. K., Zimand, E., and Calamaras, M. R. (2013). Virtual reality exposure therapy for social anxiety disorder: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, 81(5):751–760.

Andsager, J. L., Bemker, V., Choi, H. L., and Torwel, V. (2006). Perceived similarity of exemplar traits and behavior - Effects on message evaluation. *Communication Research*, 33(1):3–18.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. *Groups, Leadership, and Men. S*, pages 222–236.

Bailenson, J. N., Swinth, K., Hoyt, C., Persky, S., Dimov, A., and Blascovich, J. (2005). The Independent and Interactive Effects of Embodied-Agent Appearance and Behavior on Self-Report, Cognitive, and Behavioral Markers of Copresence in Immersive Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 14(4):379–393.

Bandura, A. (1977). *Social Learning Theory*, volume 1. Prentice Hall.

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control.* Worth Publishers.

Bandura, A. (2001). Social cognitive theory of mass communication. *Media Psychology*, 3(3):265–299.

Bandura, A. and Adams, N. E. (1977). Analysis of Self-Efficacy Theory of Behavioral Change. *Cognitive Therapy and Research*, 1:287–310.

Bandura, A., Ross, D., and Ross, S. A. (1963). Vicarious reinforcement and imitative learning. *Journal of Abnormal and Social Psychology*, 67(6):601–607.

Baron, R. A. (1988). Negative Effects of Destructive Criticism - Impact on Conflict, Self-Efficacy, and Task-Performance. *Journal of Applied Psychology*, 73:199–207.

Beidel, D. C., Turner, S. M., Jacob, R. G., and Cooley, M. R. (1989). Assessment of social phobia: Reliability of an impromptu speech task. *Journal of Anxiety Disorders*, 3(3):149–158.

Broekens, J., Qu, C., and Brinkman, W.-P. (2012). Dynamic Facial Expression of Emotion Made Easy. In *Technical report. Interactive Intelligence, Delft University of Technology.* Technical report. Interactive Intelligence, Delft University of Technology.

Brown, M. and Stopa, L. (2007). Does anticipation help or hinder performance in a subsequent speech? . *Behavioural and Cognitive Psychotherapy*, 35:133–147.

Chaiken, S. and Maheswaran, D. (1994). Heuristic processing can bias systematic processing: effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology*, 66(3):460–473.

Colman, A. M. and Olver, K. R. (1978). Reactions to Flattery as a Function of Self-Esteem - Self- Enhancement and Cognitive Consistency Theories. *British Journal of Social and Clinical Psychology*, 17(25-29).

Darley, J. M. and Latane, B. (1968). Bystander Intervention in Emergencies - Diffusion of Responsibility. *Journal of Personality and Social Psychology*, 8(4p1):377–383.

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, 7(2):117–140.

Festinger, L. (1957). *A theory of cognitive dissonance.* Peterson, Oxford: Row.

Foa, E. B. and Kozak, M. J. (1986). Emotional Processing of Fear - Exposure to Corrective Information. *Psychological Bulletin*, 99(1):20–35.

Fogg, B. J. and Nass, C. (1997). Silicon sycophants: The effects of computers that flatter. *International Journal of Human-Computer Studies*, 46(5):551–561.

Fox, J. and Bailenson, J. N. (2009). Virtual Self-Modeling: The Effects of Vicarious Reinforcement and Identification on Exercise Behaviors. *Media Psychology*, 12(1):1–25.

Geen, R. G. (1989). Alternative conceptions of social facilitation. In Paulus, P., editor, *Psychology of Group Influence Hillsdale*, pages 15–51. Lawrence Erlbaum Associates, Mahwah, NJ.

Goldstein, N. J., Cialdini, R. B., and Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conversation in hotels. *Journal of Consumer Research*, 35:472–482.

Gordon, R. A. (1996). Impact of ingratiation on judgments and evaluations: A meta-analytic investigation. *Journal of Personality and Social Psychology*, 71:54–70.

Hartanto, D., Kampmann, I. L., Morina, N., Emmelkamp, P. M. G., Neerincx, M. A., and Brinkman, W.-P. (2014). Controlling Social Stress in Virtual Reality Environments. *PLoS ONE*, 9(3):e92804.

Heider, F. (1944). Social perception and phenomenal causality. *Psychological Review*, 51:358–374.

Hope, D. A. and Heimberg, R. G. (1985). Public and private self-consciousness in a social phobic sample.

Hope, D. A., Heimberg, R. G., Zollo, L. J., Nyman, D. J., and O'Brien, G. T. (1987). Thought listing in the natural environment: Valence and focus of listed thoughts among socially anxious and nonanxious subjects.

Hunt, P. J. and Hillery, J. M. (1973). Social Facilitation in a Coaction Setting - Examination of Effects over Learning Trials. *Journal of Experimental Psychology*, 9:563–571.

Johnson, D., Gardner, J., and Wiles, J. (2004). Experience as a moderator of the media equation: the impact of flattery and praise. *International Journal of Human-Computer Studies*, 61(3):237–258.

Kashdan, T. B. and Roberts, J. E. (2004). Social Anxiety's Impact on Affect, Curiosity, and Social Self-Efficacy During a High Self-Focus Social Threat Situation. *Cognitive Therapy and Research*, 28(1):119–141.

Kozlov, M. D. and Johansen, M. K. (2010). Real Behavior in Virtual Environments: Psychology Experiments in a Simple Virtual-Reality Paradigm Using Video Games. *Cyberpsychology Behavior and Social Networking*, 13(6):711–714.

Lee, E. J. (2008). Flattery may get computers somewhere, sometimes: The moderating role of output modality, computer gender, and user gender. *International Journal of Human-Computer Studies*, 66(11):789–800.

Lee, J. (2011). Modeling side participants and bystanders: the importance of being a laugh track. In *Intelligent Virtual Agents*, pages 240–247.

Ling, Y., Brinkman, W.-P., Nefs, H. T., Qu, C., and 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., Nefs, H. T., Morina, N., Heynderickx, I., and Brinkman, W.-p. (2014). A meta-analysis on the relationship between self-reported presence and anxiety in virtual reality exposure therapy for anxiety disorders. *PLoS ONE*, 9(5):e96144.

Locke, E. A., Frederick, E., Lee, C., and Bobko, P. (1984). Effect of self-efficacy, goals, and task strategies on task performance. *Journal of Applied Psychology*, 69(2):241–251.

Meichenbaum, D. H. (1971). Examination of Model Characteristics in Reducing Avoidance Behavior. *Journal of Personality and Social Psychology*, 17(3):298–307.

Miller, C. (1984). Self-schemas, gender, and social comparison: A clarification of the related attributes hypothesis. *Journal of Personality and Social Psychology*, 46:1222–1229.

Pan, X., Gillies, M., Barker, C., Clark, D. M., and Slater, M. (2012). Socially anxious and confident men interact with a forward virtual woman: an experimental study. *PLoS ONE*, 7(4):e32931.

Park, S. and Catrambone, R. (2007). Social facilitation effects of virtual humans. *Human Factors*, 49(6):1054–1060.

Paul, G. L. (1966). *Insight Vs. Desensitization in Psychotherapy: An Experiment in Axiety Reduction.* Stanford University Press.

Pertaub, D.-P., Slater, M., and 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.

Price, M., Mehta, N., Tone, E. B., and Anderson, P. L. (2011). Does engagement with exposure yield better outcomes? Components of presence as a predictor of treatment response for virtual reality exposure therapy for social phobia. *Journal of Anxiety Disorders*, 25(6):763–770.

Qu, C., Brinkman, W.-P., Ling, Y., Wiggers, P., and Heynderickx, I. (2013). Human perception of a conversational virtual human: an empirical study on the effect of emotion and culture. *Virtual Reality*, 17(4):307–321.

Qu, C., Brinkman, W.-p., Ling, Y., Wiggers, P., and Heynderickx, I. (2014). Conversations with a virtual human: Synthetic emotions and human responses. *Computer in Human Behavior*, 34:58–68.

Reeves, B. and Nass, C. (1996). *The Media Equation.* Cambridge University Press.

Scherbaum, C. (2006). Measuring General Self-Efficacy A Comparison of Three Measures Using Item Response Theory. *Educational and Psychological Measurements*, 66(6):1047–1063.

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

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5):429–434.

Schunk, D. H. (1984). Sequential attributional feedback and children's achievement behaviors. *Journal of Educational Psychology*, 76:1159–1169.

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., Pertaub, D.-P., and Steed, A. (1999). Public speaking in virtual reality: facing an audience of avatars. *IEEE Computer Graphics and Applications*, 19(2):6–9.

Slater, M., Rovira, A., Southern, R., Swapp, D., Zhang, J. J., Campbell, C., and Levine, M. (2013). Bystander responses to a violent incident in an immersive virtual environment. *PLoS ONE*, 8(1):e52766.

Somerville, L. H., Heatherton, T. F., and Kelley, W. M. (2006). Anterior cingulate cortex responds differentially to expectancy violation and social rejection. *Nature neuroscience*, 9(8):1007–1008.

Takata, C. and Takata, T. (1976). The influence of models in the evaluation of ability. *Japanese Journal of Psychology*, 47:74–84.

Thomasson, P. and Psouni, E. (2010). Social anxiety and related social impairment are linked to self-efficacy and dysfunctional coping. *Scandinavian Journal of Psychology*, 51(2):171–178.

Vassilopoulos, S. P. (2008). Coping strategies and anticipatory processing in high and low socially anxious individuals. *Journal of Anxiety Disorders*, 22(1):98–107.

Villani, D., Repetto, C., Cipresso, P., and Riva, G. (2012). May I experience more presence in doing the same thing in virtual reality than in reality? An answer from a simulated job interview. *Interacting with Computers*, 24(4):265–272.

Vonk, R. (2002). Self-serving interpretations of flattery: Why ingratiation works. *Journal of Personality and Social Psychology*, 82:515–526.

Walster, E. and Festinger, L. (1962). The effectiveness of "overheard" persuasive communications. *Journal of Personality and Social Psychology*, 65:395–402.

Wetzel, C. G. and Insko, C. A. (1982). The similarity-attraction relationship: Is there an ideal one? *Journal of Experimental Social Psychology*, 18(3):253–276.

Wolpe, J. (1969). *The Practice of Behavior Therapy.* Pergamon Press, New York.

Zitek, E. M. and Hebl, M. R. (2007). The role of social norm clarity in the influenced expression of prejudice over time. *Journal of Experimental Social Psychology*, 43(6):867–876.

Chapter 6

Conclusion and Discussion

This thesis investigated how to adapt a virtual environment to affect human experience during a conversation with a virtual human within the domain of virtual reality exposure therapy for treating people with a social anxiety disorder. The aspects of the virtual environment considered in this thesis include priming the communication with media material in the surrounding, the virtual dialog partner and the virtual bystanders. The study was set out to answer the following main research question:

Can and in what way do the virtual surrounding, the behaviour of a virtual dialog partner, and the behaviour of the virtual bystanders have an effect on an individual who is engaged in a conversation with a virtual dialog partner?

Four hypotheses were formulated to answer the main research question:

1. Priming pictures and videos increase the chance that individuals use specific keywords in their answers when having a conversation with a virtual human.

2. The virtual human's expressed valence is perceived as more intense in the speaking phase than in the listening phase.

3. By expressing a positive or negative emotion, a virtual human can elicit a corresponding emotional state in a human conversation partner and affect the satisfaction towards the conversation.

4. Virtual bystanders can affect a person's beliefs and behaviour during a virtual conversation.

The results presented in this thesis demonstrate that the virtual surrounding, the behaviour of a virtual conversation partner, and the behaviour of virtual bystanders indeed all affect the individual during a conversation with a virtual human; hence answering the ‘can’ part of the main research question. The ‘how’ part of the research question may be answered by addressing the four hypotheses, as each gives insight in how an individual may be affected in a virtual conversation. Thus, the conclusions in this thesis are structured by summarizing the arguments for these four hypotheses.

The effect of priming pictures and videos

Support for the first hypothesis was established through two empirical studies, investigating the priming effect of media material such as pictures and videos, in a conversation scenario, first in the real word and then in the virtual world. Twenty participants were recruited for the first study; they watched videos, and then were exposed to pictures while answering several open questions from the experimenter. The second study shared the same experimental setting as the first one, except that it was carried out in virtual reality. Again, twenty participants were recruited and they were asked to answer a number of open questions this time posed by a virtual human. The results of both studies showed that participants tend to mention the target keywords when exposed to priming videos and pictures, indicating that both priming materials - pictures and videos - had a guiding effect on the conversation topic of humans. The participants even forgot to mention the content they typically would mention when no priming pictures were shown in the environment. These results therefore demonstrate that virtual surroundings can prime people and affect what people say in a conversation with real human or virtual human.

The difference in human valence perception between speaking and listening

The second hypothesis was also empirically examined. The emotion of a virtual human was manipulated separately during the listening phase and the speaking phase, and 24 recruited participants were allocated into two equally sized groups according to their nationality, i.e., a Chinese and non-Chinese group. The participants were asked to rate how they perceived the emotional valence of a virtual Chinese lady. The results showed that valence expression during the speaking phase was perceived as more intense compared to valence expression during the listening phase, which supports hypothesis 2. Both Chinese and non-Chinese participants could perceive the valence of the emotion of the virtual Chinese human, and no significant difference between these two groups was found. Instead, the ratings of the Chinese and non-Chinese groups were highly correlated.

The effect of the behaviour of a virtual conversation partner

The third hypothesis was supported by the findings of another empirical experiment using the same virtual Chinese lady as in the previous experiment. This virtual lady now expressed positive, neutral or negative emotions during either the listening or speaking phase of a conversation, and 24 Chinese participants were requested to score the their emotion and their satisfaction with the dialog. The analyses of the data suggested that positive compared to negative synthetic emotions expressed by a talking virtual human can elicit a more positive emotional state in a person, and can create more satisfaction towards the conversation. A larger effect on reported valence and discussion satisfaction was found for the synthetic emotions manipulated in the speaking phase compared to the listening phase of the virtual lady. Actually, support for the third hypothesis was only found in the speaking phase of the virtual human, while the differences found in the listening phase were not significant. The explanation may be twofold: besides the additional verbal channel to express emotions in the speaking phase, the participants might also have spent less attention to the virtual human when they were talking and the virtual human was listening. We also found that participants with more speaking confidence reported a more positive emotional state and more satisfaction with the discussion than participants with less speaking confidence. The results also showed that a virtual human expressing randomly positive or negative emotion had a negative effect on discussion satisfaction compared to a virtual human expressing a neutral emotional state.

The effect of the behaviour of virtual bystanders

Support for the last hypothesis was established also by an experiment in which the attitude expressed by virtual bystanders was manipulated to be either positive or negative towards the virtual peer speakers and towards the participants during a virtual English lesson. The bystanders affected the participants' belief and behaviour in several ways. For example, the results showed that the participants established more self-efficacy and showed less avoidance behaviour when the bystanders expressed a positive attitude compared to a negative attitude towards them. Also witnessing bystanders commenting negatively on the other students' performance resulted in higher heart rate when it was the participant's turn to speak. The results also showed that the bystanders' positive attitude towards both the virtual peer speakers and the participants evoked the lowest level of self-reported anxiety in the participants. Besides these effects, the experiment also showed the impact of consistency versus inconstancy in bystanders' behaviour. Bystanders' inconsistent attitude towards the virtual peer speakers and the participants resulted in a larger change in participants' rating of their own performances and their belief of the virtual teacher's satisfaction with their performance. The experiment also demonstrated the possibility to replicate the effect of flattering in virtual reality. When the virtual bystanders made positive comments, instead of negative about the participants, participants liked the bystanders more, rated their performance higher, and found them more supportive.

6.1 Limitations

To appreciate the work presented in this thesis, it is also important to consider its limitations. First, although the study focused on social anxiety disorder, only participants from a technical university were recruited and none of these participants suffered from social anxiety disorder. Socially phobic people may be even more sensitive to socially important cues, such as whether the virtual human looks at them, or whether the virtual human shows a positive attitude towards them (Clark and McManus, 2002). Still in this study, the level of speaking confidence was found to relate to the emotional experience during the virtual conversation. Less confident participants reported a more negative emotional state and were less satisfied with the discussion than participants with more speaking confidence. Recently Hartanto et al. (2014) also reported similar findings with virtual conversations. Therefore, although promising, additional research with a group of socially phobic people is needed before making any firm claim about the generalization of the thesis' findings to this group.

Second, the second and third study used a Chinese lady as the virtual conversation partner, and possible cultural differences were only tested between Chinese and non-Chinese participants. Also because of the language used by the virtual human, only Chinese participants were recruited to test the third hypothesis addressing the impact of the virtual lady's synthetic emotions on participants' behaviour. Therefore, using only a virtual Chinese lady may limit the generalisation of our findings to other cultures. On the other hand, the study found a strong correlation between Chinese and no Chinese participants for their perception of the valence of the emotion expressed by the virtual human. Hence, it seems that emotions are similarly recognized by different cultures, but still the impact of the perceived emotions on experience and dialog satisfaction may be different between different cultures.

Third, in the experiment evaluating the behaviour of virtual bystanders, neither the attitude of the virtual bystanders nor the response of the virtual teacher changed as a reaction upon the participants' performance. Still this study and Hartanto et al. (2014) have shown that virtual humans who provide positive or negative feedback in a dialog may affect individuals' emotion and behaviour. Therefore, making the virtual bystanders or the virtual teacher show their attitude as a response to the participants' performance may affect the participants' belief and behaviour in the virtual environment even more, as they may experience that their effort, i.e., trying to perform better, may have an impact on their environment.

Fourth, the emotion expressed by the virtual humans, i.e., the virtual Chinese lady and the virtual bystanders only included a limited set of facial expressions (such as anger and happiness) of a much larger set of possible emotions. Emotions such as sadness, fear and frustration may also make the individuals have the impression that the virtual human evaluates them negatively. Whether using the latter emotions instead of the ones we used in our studies would have made our conclusions stronger or less strong is unclear at this moment, and requires additional research.

6.2 Contributions

6.2.1 Scientific contributions

The main contributions of the work presented in this thesis are insights provided into how aspects of a virtual environment may affect humans and how humans react to these aspects when they are engaged in a conversation with a virtual human. The aspects investigated are: (1) the surrounding of a virtual human and how that surrounding may prime humans to use specific keywords in their communication, (2) the virtual conversation partner and its ability to elicit emotional reactions through the emotions it expresses, and (3) the virtual bystanders and their ability to elicit emotional reactions and affect beliefs through the attitude they express. The thesis demonstrates that phenomena known from the real world, such as priming, vicarious experience, and universality of facially expressed emotions, can be replicated in virtual reality. Reversely, observations made in the virtual world may also be generalised to the real world. As others (Fox and Bailenson, 2009; Kozlov and Johansen, 2010; Park and Catrambone, 2007; Slater et al., 2013) also have argued, virtual reality offers a very controlled way to study real world phenomena. In this context, this thesis contributes to understanding human-human conversation, for example with respect to the impact of emotions expressed in the listening or speaking phase, and with respect to the impact the attitude of bystanders has on self-efficacy, attitude towards others, and beliefs about the social world.

Although the work presented focused on conversations within the setting of VRET for the treatment of social anxiety disorder, the findings may have wider implications for virtual conversations in general. Specifically the findings of this thesis on people's reactions towards synthetic emotions during a conversation and also the effect of bystanders seem not limited to anxiety provoking situations for patients, but may be applicable to the public in general, especially since the participants of our studies were drawn from a non-clinical population. Furthermore, the priming phenomenon observed in this thesis adds to work about priming mechanisms in virtual environments. For example, Anderson and Dill (2000) demonstrated that playing aggressive video games primes aggressive thoughts and behaviours. Pena and Blackburn (2013) studied the priming effect of e.g., a virtual library or a virtual caf on interpersonal perceptions and behaviour. Whereas in the latter studies priming involved the entire virtual environment, this thesis shows that specific elements in a virtual environment such as pictures, may also prime specific behaviour.

6.2.2 Practical contributions

Development and use of VRET

Using priming videos and pictures in virtual environments may help to improve the perceived naturalness of a virtual conversation as the computer is more able to give suitable verbal replies. This mechanism aims at improving the level of perceived presence, in this case the feeling of being engaged in a conversation. The feeling of presence is a key element for VRET as Ling et al. (2014) in a recent meta-analysis found a positive correlation between self-reported presence and anxiety.

The work presented in the thesis also points to a number of ways to control anxiety-evoking stimuli in a VRET system for treating social phobia. For example, manipulating the social interaction individuals observe prior to being exposed to the same social interaction gives therapists a mean to control anticipation anxiety. During the exposure, therapists may control anxiety by manipulating the attitude expressed by bystanders and by the conversation partner. Finally, exposing individuals randomly to positive or negative emotions may also negatively affect the experience of the conversation. This finding implies that even without intentionally using random emotional shifts, therapists should be aware that if they switch between positive and negative emotions too often, they may elicit the same negative effect.

Besides having the ability to evoke anxiety, VRET systems may also support therapists in measuring anxiety. Besides self-reported and physiological measures used in VRET for treating other anxiety disorders, speech length was found to be a useful avoidance measure in a social anxiety setting. Therefore, using speech detection technology may extend a VRET system with an unobtrusive measure.

Application Designers

The work also has some implications for application designers. For example, we found no difference between Chinese and non-Chinese participants in their perception of the valence of emotions expressed by a virtual human, which supports the idea of universality of facial expressions of emotion (Ekman, 1994; Matsumoto, 2007). The high consistency in the evaluation of emotions of virtual humans between Chinese and non-Chinese people, questions the need for tailored made virtual humans which target different cultural groups instead of targeting a multi-cultural consumer group.

Application designers that want to use virtual conversations to elicit emotions, may focus specifically on the speaking phase instead of the listening phase of a virtual human, as the work presented in this thesis found the speaking phase to be more effective in influencing people's perception and emotion.

As mentioned before, simply giving a virtual human random emotions may reduce the satisfaction a person experiences when talking with such a virtual human. Therefore, designers should consider giving a virtual human a neutral expression or implement conversation appropriate emotions.

Finally, the implication of the priming effect for designers means that they should carefully consider the design of virtual environments if they want to design virtual conversations, since our findings showed that pictures may have the ability to overwrite verbal response that people typically would give.

Teachers

The experiment measuring the impact of the behaviour of bystanders gives insight into short-term effects of experiencing bullying in a classroom. The results demonstrated that negative whispers among fellow classmates may affect someone's beliefs about his or her own self-efficacy, raise anxiety, and also result in negative beliefs about fellow classmates and the teacher. In addition, the study also gives insight into the victims' beliefs about a teacher who is reluctant to intervene. In the absence of a teacher that intervenes, victims believe that the teacher likes them less and are less pleased with their own performance. This suggests that inaction of the side of the teacher is interpreted as agreeing with the students.

As bullying may have long-term effects on the mental health of both the bullies and the victims (Swearer et al., 2001), increasingly more attention is devoted to developing strategies to avoid or intervene with bullying (Howard et al., 2001; Sanchez et al., 2001). Therefore, our results may help teachers to realise that positive expressed attitudes among fellow classmates have a positive effect on participants' beliefs and emotions. Hence, teachers may therefore try to create a positive climate in the classroom.

6.3 Future work

The thesis also provides suggestions for future research. First, the work on priming may be extended by examining the potential of dynamic priming to control the flow of a conversation. Instead of static pictures in a virtual world, exposing a person to dynamically generated priming stimuli during the conversation may influence the flow of the dialog. The latter could be done for example by having a TV screen in the virtual world, on which specific priming material embedded in a TV program is shown. For example, Isnanda et al. (2014) used TV news flashes to elicit paranoid thoughts in a virtual restaurant environment. As a consequence, with dynamic priming it may be possible to repeatedly expose a patient to the same VR world, though each time having another conversation with a virtual human. Future work may also compare the priming effect between different priming materials, such as pictures, videos and virtual objects e.g., furniture or decoration, and find the best way of priming people during virtual reality exposure.

Second, the similarity between Chinese and non-Chinese people in their perception of the valence of emotions expressed by virtual humans justifies extending this work to people with other cultures, to virtual humans with other ethic appearances, and to other emotions.

Third, as this work already demonstrated that emotions expressed mainly by the upper body can affect human perception, emotion, attitude, beliefs, and behaviour, exploring full body emotion expression seems the next step. Literature already provides evidence that full body emotion expression does indeed play an important role in the expression and perception of emotion or attitude (Gross et al., 2010; Kang et al., 2014; Kleinsmith and Bianchi-Berthouze, 2013). Therefore, applying full body emotion expression in the setting of a conversation with a virtual human could potentially enrich the human experience.

Fourth, the emotions expressed by the virtual human and the virtual bystanders only included a limited set of facial expressions such as anger and happiness. Future research could investigate whether emotions such as sadness, fear and frustration expressed by the virtual humans could also make the individuals have the impression that they were negatively evaluated.

Fifth, the experiment investigating the impact of the behaviour of virtual bystanders demonstrated that people's experience in virtual reality may affect their self-efficacy. As self-efficacy is regarded as a predictor of actual behaviour (Bandura and Adams, 1977; Locke et al., 1984), future virtual reality studies may consider collecting data about people's self-efficacy as these data would give insight into how their behaviour in the real world is changed by having experienced exposure to the virtual world.

6.4 Final remark

This thesis aims to understand the effect of three aspect of a virtual environment, i.e., the surrounding of a virtual human, the virtual conversational partner and the virtual bystanders, on an individual's experience and behaviour during a virtual conversation. As such, it is possible to control an individual's experience in a virtual conversation, and to provide therapists with more options to control stimuli in virtual reality exposure therapy systems for treating people with social phobia. Four main empirical studies were conducted to test the hypotheses of this thesis. The results showed that priming materials such as videos and pictures have a guiding effect on humans when they have a conversation with a virtual human. This shows the possibility to improve the dialog manager of the system by showing topic-related virtual materials in order to unobtrusively guide patients. Emotions expressed in the speaking phase by a virtual conversation partner were perceived as more intense than emotions expressed in the listening phase, and they had a larger effect on people's valence and how they experienced the discussion. A positive instead of negative attitude of a virtual conversation partner elicited a more positive emotional state in humans as well. Similarly, a positive attitude of virtual bystanders towards a person elicited higher self-efficacy and people showed less avoidance behaviour as compared to a negative attitude of the bystanders.

In short, by manipulating virtual objects, virtual bystanders, or the virtual conversation partner a therapist may affect the behaviour, emotions, and beliefs of a person. Although the empirical studies were conducted in the context of social anxiety disorder, the findings with respect to the three aspects may also apply to other virtual reality application domains such as education, gaming and virtual coaching.

Bibliography

Anderson, C. A. and Dill, K. E. (2000). Video Games and Aggressive Thoughts, Feelings, and Behavior in the Laboratory and in Life. *Journal of Personality and Social Psychology*, 78(4):772–790.

Bandura, A. and Adams, N. E. (1977). Analysis of Self-Efficacy Theory of Behavioral Change. *Cognitive Therapy and Research*, 1:287–310.

Clark, D. M. and McManus, F. (2002). Information processing in social phobia. *Biological Psychiatry*, 51(1):92–100.

Fox, J. and Bailenson, J. N. (2009). Virtual Self-Modeling: The Effects of Vicarious Reinforcement and Identification on Exercise Behaviors. *Media Psychology*, 12(1):1–25.

Gross, M. M., Crane, E. A., and Fredrickson, B. L. (2010). Methodology for Assessing Bodily Expression of Emotion. *Journal of Nonverbal Behavior*, 34(4):223–248.

Hartanto, D., Kampmann, I. L., Morina, N., Emmelkamp, P. M. G., Neerincx, M. A., and Brinkman, W.-P. (2014). Controlling Social Stress in Virtual Reality Environments. *PLoS ONE*, 9(3):e92804.

Howard, N. M., Horne, A. M., and Jolliff, D. (2001). Self-Efficacy in a New Training Model for the Prevention of Bullying in Schools. *Journal of Emotional Abuse*, 2(2-3):181–191.

Isnanda, R. G., Brinkman, W.-P., Veling, W., van der Gaag, M., and Neerincx, M. (2014). Controlling a Stream of Paranoia Evoking Events in a Virtual Reality Environment (submitted). *Annual Review of Cybertherapy and Telemedicine.*

Kang, N., Brinkman, W.-P., van Riemsdijk, M. B., and Neerincx, M. A. (2014). An Expressive Virtual Audience with Flexible Behavioral Styles. *IEEE Transactions on Affective Computing.*

Kleinsmith, A. and Bianchi-Berthouze, N. (2013). Affective Body Expression Perception and Recognition: A Survey. *IEEE Transactions on Affective Computing*, 4(1):15–33.

Kozlov, M. D. and Johansen, M. K. (2010). Real Behavior in Virtual Environments: Psychology Experiments in a Simple Virtual-Reality Paradigm Using Video Games. *Cyberpsychology Behavior and Social Networking*, 13(6):711–714.

Ling, Y., Nefs, H. T., Morina, N., Heynderickx, I., and Brinkman, W.-p. (2014). A meta-analysis on the relationship between self-reported presence and anxiety in virtual reality exposure therapy for anxiety disorders. *PLoS ONE*, 9(5):e96144.

Locke, E. A., Frederick, E., Lee, C., and Bobko, P. (1984). Effect of self-efficacy, goals, and task strategies on task performance. *Journal of Applied Psychology*, 69(2):241–251.

Park, S. and Catrambone, R. (2007). Social facilitation effects of virtual humans. *Human Factors*, 49(6):1054–1060.

Pena, J. and Blackburn, K. (2013). The Priming Effects of Virtual Environments on Interpersonal Perceptions and Behaviors. *Journal of Communication*, 63(4):703–720.

Sanchez, E., Robertson, T. R., Lewis, C. M., Rosenbluth, B., Bohman, T., and Casey, D. M. (2001). Preventing bullying and sexual harassment in elementary schools: The Expect Respect Model. *Journal of Emotional Abuse*, 2(2-3):157–180.

Slater, M., Rovira, A., Southern, R., Swapp, D., Zhang, J. J., Campbell, C., and Levine, M. (2013). Bystander responses to a violent incident in an immersive virtual environment. *PLoS ONE*, 8(1):e52766.

Swearer, S. M., Song, S. Y., Cary, P. T., Eagler, J. W., and Mickelson, W. T. (2001). Psychosocial Correlates in Bullying and Victimization: The Relationship Between Depression, Anxiety, and Bully/Victim Status. *Journal of Emotional Abuse*, 2(2-3):95–121.

Acknowledgement

First and foremost I offer my sincerest gratitude to my promoter Ingrid Heyndrickx and my daily supervisors Willem-Paul Brinkman and Pascal Wiggers. You supported me through my PhD life with patience and knowledge whilst allowing me the room to work in my own way. I attribute the level of my PhD degree to your encouragement and effort. Without you my thesis would not have been completed.

I also would like to thank Harold Nefs and Joost Broekens. Although you are not my daily supervisors, your strong theoretical knowledge and broad academic vision helped me greatly throughout my PhD career. All my participants, thank you for helping me with my experiments so I can draw all the promising conclusions. Judith A. Redi, Harold Nefs, Reyhan Aydogan, Maaike Harbers, Dylan Schouten, Vanessa Vakili, Wenxin Wang and Ni Kang, thank you for helping me practice my thesis defense, all your comments and suggestions are very valuable for me. Corine Horsch and Dwi Hartanto, thank you for being my paranymphs.

Members of VRET group: Dwi Hartanto, Ni Kang, Dylan Schouten, Corine Horsch, Iris Cohen, Wenxin Wang, Vanessa Vakili and Myrthe Tielman, and former VRET members: Guntur Sandino, Niels ter Heijden, Daniel de Vliegher and Charles van der Mast, thank you all for sharing knowledge and information during the VRET meeting every week and helping me solving problems every now and then.

Only focusing on research is not enough for a PhD student. There are always trivial but inevitable things need to be dealt with. Personally I lack the patience to handle those little affairs but thankfully, our secretary Anita Hoogmoed and Onno de Wit, our technical support Ruud de Jong and Bart Vastenhouw can always solve all kinds of problems for me. Many thanks for your help.

Catholijn Jonker, thank you for leading the ii group and providing us opportunity to do research, also thank you for all the board game nights. All ii group members, thank you all for everything happened in the group during the past 4 years.

Hani Alers, Nike Gunawan, Hantao Liu and Zhenke Yang, although you left the group before me, I still would like to thank you all for those party nights we enjoyed together. My Chinese friends, Yangyang Shi, Junchao Xu, Ni Kang, Tingting Zhang, Lin Xia, Changyun Wei, Chang Wang, Zhenzhong Sun, Hao Cui, Zhenzhong Sun, Liangyue Ji, Hui Tang, Yue Shi, Xin Wang, Zhihao Zhu, Xiaofan Sun, Haiyan Wang, etc., also thank you all for countless dinners and various weekend activities we had together.

Olaf Tettero, my current group leader, thank you for accepting me to the PED group, I really enjoy my life in the group. All PED group members, thank you for all the help I got in the past 9 months. Shiqi Li and Chang Shu, thank you for letting me join your Aula group, always enjoy chatting with you during lunch time.

I thank my parents and my family for supporting me to start and finish my PhD abroad. Mom, thank you for skyping with me every weekends. Dad, thank you for all the nice ideas you shared with me. My uncles and aunts, your encouragement helped me a lot during my hazy days. My brothers, your visit in the Netherlands makes me feel home.

Finally, the special thanks are for my beloved wife Yun Ling. I really appreciate your accompany. What you did for me is too much to be listed here, so I decide to leave them out and keep them in my heart, forever.

Curriculum Vitae

Chao Qu was born in Suzhou, China on December 12th, 1983.

From 1999 to 2002, he studied in Suzhou High School of Jiangsu Province. Since then he had already had a great interest in computer science and programming. He won the first price in the computer program designing competition of Jiangsu Province.

From 2002 to 2006, he studied in Northeast Normal University, School of Physics. His major was electronic science and technology. He was the monitor of the class and the president of computer association. He received scholarships from the University every year and got his bachelor's degree in 2006.

From 2006 to 2009, he studied in Southeast University, Department of Electronic Science and Engineering. His major was image and video processing. He also did a part time job in Nanjing Huaxian Technological Co., Ltd. (http://www.smpdp.com), using FPGA/CPLD to design the driver for plasma display panel. He organized the first CPLD designing competition of Southeast University and he have a Chinese patent 'CN201177909 A Programmable Application Development Platform based on CPLD.' He obtained his master's degree in 2009.

Since the end of 2009, he is working as a PhD researcher in Delft University of Technology, the Netherlands. His research topic is life-like multi-modal avatar. To solve the problem that users always complain about the unrealistic virtual worlds, he focused on how to take the advantage of the virtual reality to make up for its own disadvantage. His research improves the user experience between human-avatar conversations and gives the user a high presence of being in the virtual reality. His research has been applied to virtual reality exposure therapy for treating all kinds of anxiety disorders such as social anxiety disorders. he gave lectures on how to make emotional avatars to masters/PhD students and employees in virtual reality companies, e.g., CleVR (http://clevr.net) and ELXR (http://www.elxr.net). He created a lot of virtual humans and virtual reality applications for both his own research purpose and his colleagues' research.