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

Conversations with a virtual human:

Synthetic emotions and human responses

Chao Qu[1]

Delft University of Technology, Mekelweg 4, 2628 CD Delft, the Netherlands

Willem-Paul Brinkman

Delft University of Technology, Mekelweg 4, 2628 CD Delft, the Netherlands

Yun Ling

Delft University of Technology, Mekelweg 4, 2628 CD Delft, the Netherlands

Pascal Wiggers

CREATE-IT Applied Research, Amsterdam University of Applied Sciences,

Duivendrechtsekade 36-38, 1096 AH Amsterdam, the Netherlands

Ingrid Heynderickx

Human Technology Interaction Group, Eindhoven University of Technology, Eindhoven, the Netherlands and Visual Experiences Group, Philips Research Eindhoven, the Netherlands

[1] Correspondence to aquchaos@gmail.com

Tel.:+31681808384

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

Abstract

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 dialogue. 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 impact 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.

Keywords: virtual reality, virtual human, emotion, dialogue experience, social anxiety

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, Harbers, et al., 2012; Core, Traum, Lane, & Swartout, 2006), communication (Lok, 2006), interview (Link, Armsby, Hubal, & Guinn, 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, Jacobs, & Rothbaum, 2004; Anderson, Rothbaum, & Hodges, 2001; Krijn, Emmelkamp, Olafsson, & Biemond, 2004; 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 & Tarrier, 2007; Opris et al., 2012; Parsons & Rizzo, 2008; Powers & 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, van der Mast, & de Vliegher, 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, van der Mast, Sandino, Gunawan, & Emmelkamp, 2010; Gunawan, van der Mast, Neerincx, Emmelkamp, & Krijn, 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, Emmelkamp, Biemond, et al., 2004). For patients with social anxiety, the therapist also needs access to these controls (Clark & Beck, 2011) and needs more flexibility (Lanyi, Stark, Kamson, & Geiszt, 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, Alan, & Czanner, 2009; ter Heijden & Brinkman, 2011).

This paper, therefore, studies dialogue 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 dialogue.

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, Bailenson, and Rickertsen (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, Lin, Steed, Swapp, & Slater, 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 & 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).

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, Slater, and Barker (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 researchers have also studied the impact of positive behaviour of a virtual human on actual humans. De Melo, Carnevale, and Gratch (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, Brinkman, Nefs, Qu, & Heynderickx, 2012; Pertaub et al., 2002; Wong & McGee, 2012) or where a virtual human speak to a human (Baylor, Ryu, & Shen, 2003; Konstantinidis, Hitoglou-Antoniadou, Luneski, Bamidis, & Nikolaidou, 2009; Qiu & 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 a virtual human towards its 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.

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, Hattangadi, Meziane, and Pul (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.

2.3. Low Anxiety Group versus High Anxiety Group

If the dialogue 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, Pertaub, Barker, and Clark (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, Gillies, Barker, Clark, & Slater, 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.

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 (D. 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, Adam, Van Kleef, & Galinsky, 2013) and to make greater concessions (Van Kleef & 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.

3. Method

A within-subjects experiment with six conditions (see Table 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-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 1) condition and a random (indicated by r in Table 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.

Table 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

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.

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, Kramer, Gratch, and Kang (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 & 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, Qu, & Brinkman, 2012). Figure 1 shows the virtual human expressing emotions from neutral (b) to negative (a) or positive (c).

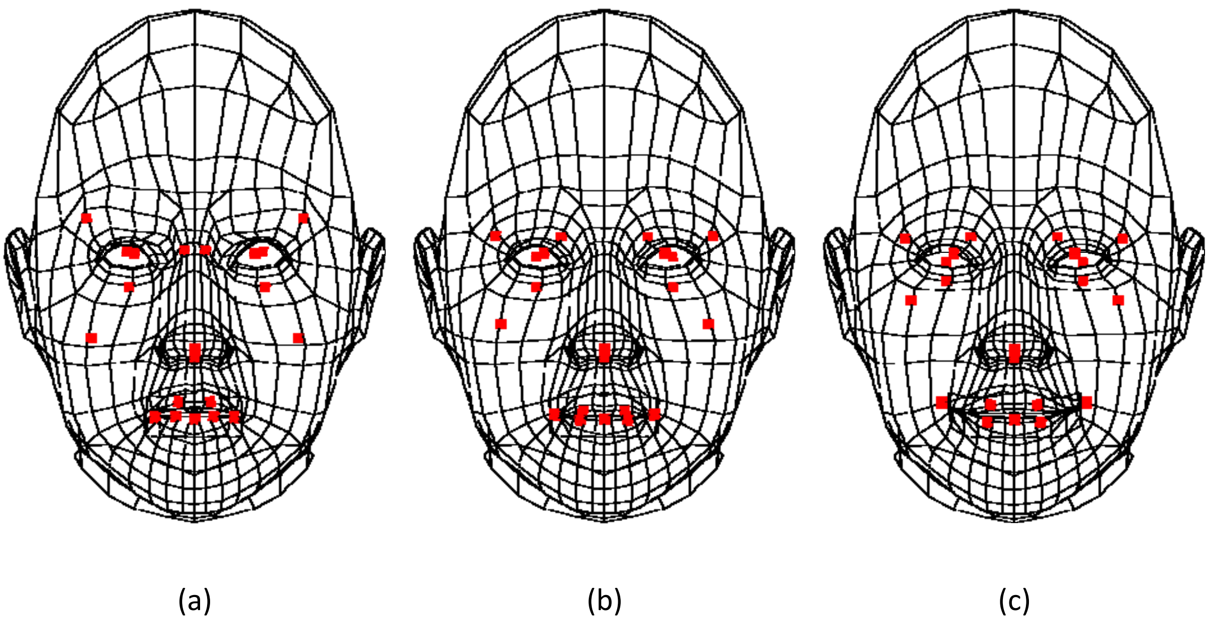

Figure 1: Emotions expressed by moving action units (i.e., small squares in the picture) attached to a face mesh: (a) angry, (b) neutral, and (c) happy

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 2c). 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 2a). The intensity of both the positive and negative emotional expressions was evaluated in a previous study (Broekens, Qu, et al., 2012) to ensure that they both could be identified by individuals. For the neutral condition, a neutral facial expression[2] was used and the lady kept looking at the participants with some slight eye and head movements (Figure 2b). 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.

[2] 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, 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 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 3a) and a happy facial expression was shown in the positive condition (Figure 3c). 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 3b). Again, the random condition existed of the combination of positive and negative emotions, controlled by a random coefficient.

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

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 dialogue tool Editor3 (ter Heijden & Brinkman, 2011; ter Heijden, Qu, Wiggers, & Brinkman, 2010) was used to create six dialogues on the following topics: research project, food, movie, China, travelling, and living in the Netherlands. Each dialogue 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 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.

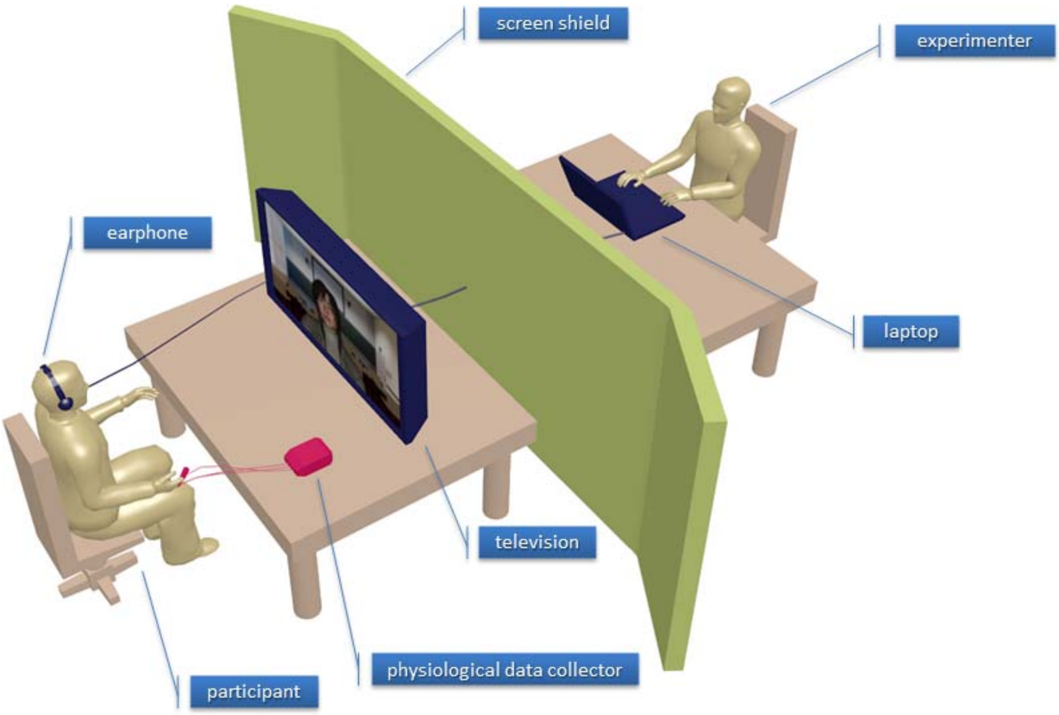

Figure 4 : Experimental Setup

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 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.* = .711), and also no significant two-way interaction effect was found ($F(1, 11) = 0.44$, *p.* = .522). 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.

3.4. Measurements

3.4.1. 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, Jones, Rieger, and Snell (1997) showed that the PRCS index did not differ across age and gender.

3.4.2. Dialogue Satisfaction

The Dialogue Experience Questionnaire (DEQ) (ter Heijden & 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., dialogue 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.

3.4.3. 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, Bradley, & Cuthbert, 1997; J. D. Morris, 1995), sounds (Bradley & Lang, 2007), robot gesture expression

(Haring, Bee, & Andre, 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).

3.4.4. Presence questionnaire

Participants were asked to complete the Igroup Presence Questionnaire (IPQ) (Schubert, Friedmann, & Regenbrecht, 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.

3.4.5. Dialogue 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, Turner, Jacob, & Cooley, 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.

3.4.6. 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). 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.

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, 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. 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[3]. The alpha level was set at .05 for all the tests.

As some of the dependent 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, Findlater, Gergle, & Higgins, 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[4] for non-stereoscopic monitor and the median IPQ score in the *L-S-* (Mdn = 41, IQR = 13, z = 0.53, p = .596), *L0S0* (Mdn = 45, IQR = 16.75, z = 1.87, p = .061) and *LrSr* (Mdn = 42.5, IQR = 16.25, z = 1.49, p = .136) 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.

4.1. 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 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-*.

---

[3] 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.

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

4.1.1. Dialogue 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 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.

Table 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	$F(1,49) = 0.03$, $p = .865$

4.1.2. 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 3. The results of the linear-mixed-model analysis on the aligned ranks data (see Table 4) show that synthetic emotions in the speaking phase affected the participants' valence and dominance significantly. Participants reported a 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 5a. Table 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 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.

Table 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: Results of Mixed-effect Model Analysis of Variance for the SAM scores.

	Valence	Arousal	Dominance
PRCS	$F(1,18) = 9.05$, $p = .008$	$F(1,17) = 1.45$, $p = .245$	$F(1,17) = 1.43$, $p = .248$
Listening	$F(1,45) = 1.50$, $p = .227$	$F(1,46) = 2.80$, $p = .101$	$F(1,50) = 0.04$, $p = .850$
Speaking	$F(1,47) = 21.0$, $p < .001$	$F(1,44) = 0.01$, $p = .906$	$F(1,50) = 5.69$, $p = .021$
PRCS× Listening	$F(1,45) = 0.19$, $p = .665$	$F(1,47) = 8.06$, $p = .007$	$F(1,50) < 0.01$, $p = .981$
PRCS× Speaking	$F(1,46) = 0.25$, $p = .620$	$F(1,45) = 2.34$, $p = .133$	$F(1,50) = 0.05$, $p = .829$
Listening× Speaking	$F(1,45) = 0.38$, $p = .540$	$F(1,44) = 0.57$, $p = .456$	$F(1,50) = 0.23$, $p = .632$
PRCS×Listening ×Speaking	$F(1,45) = 0.95$, $p = .334$	$F(1,44) = 0.41$, $p = .525$	$F(1,50) < 0.01$, $p = .948$

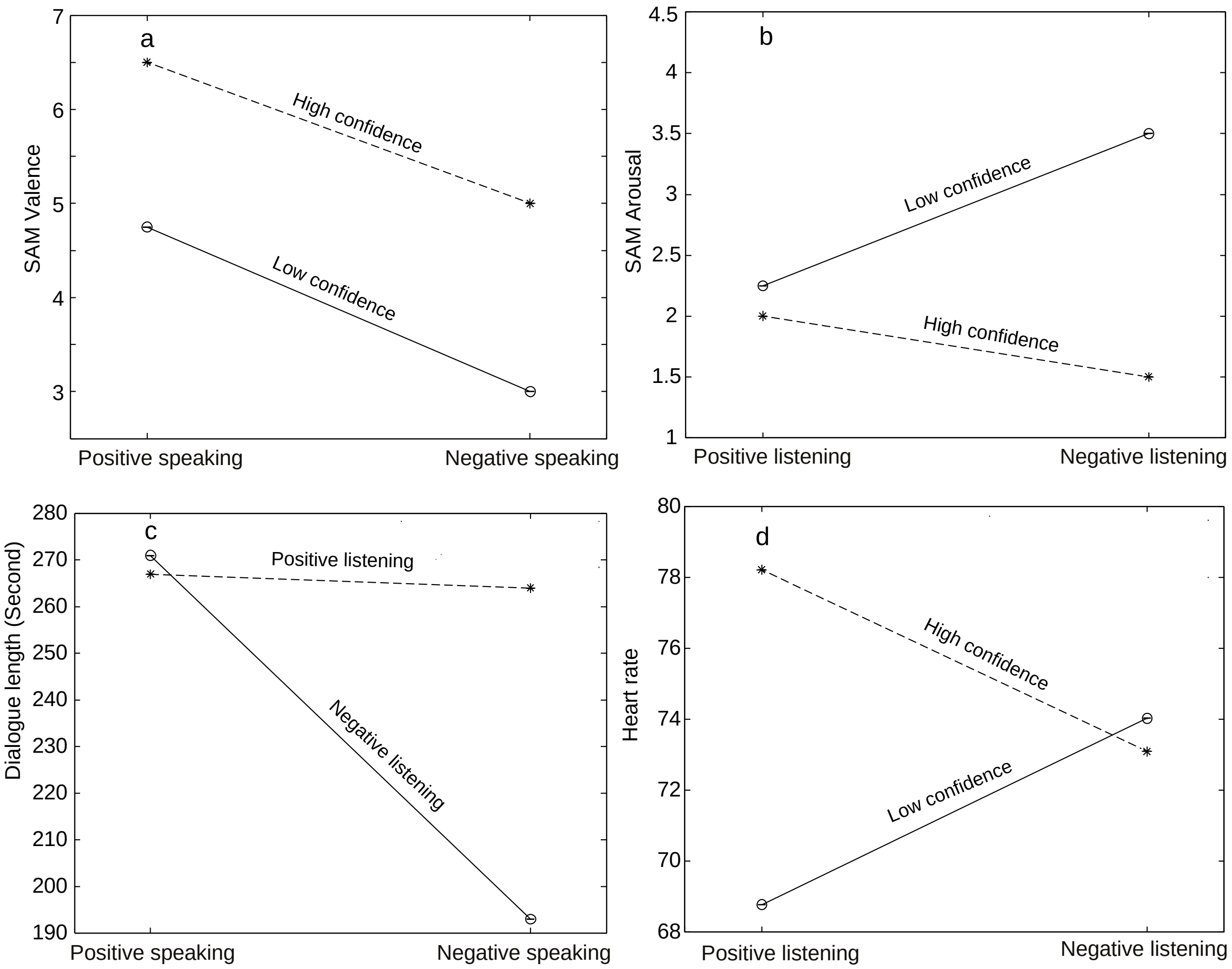

Figure 5: Median of SAM valence score (a), SAM arousal score (b), participants' dialogue length (c), and heart rate (d).

4.1.3. Dialogue 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 267.0 (182.8), 264.0 (214.0), 271.0 (231.8), and 193.0 (164.0) respectively (see also Figure 5c). Table 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 5 shows no significant main effect of the synthetic emotions in the listening phase on dialogue length. In addition, Table 5 shows a significant interaction between the emotions expressed in the speaking and listening phase. As can be seen in Figure 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 5: Results of Mixed-effect Models Analysis of Variance for the dialogue length

	Dialogue length[5]
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	$F(1,41) = 0.02$, $p = .898$

4.1.4. 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 6 shows a significant interaction between the PRCS groups and the listening behaviour on the heart rate data. As can be seen in Figure 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

[5] The interaction effect of listening and speaking on dialogue length was not significant ($F(1,69) = 2.41$, p = .125) when the analysis was conducted using PRCS between-subjects variable with three levels.

from positive to negative. Table 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.

Table 6: Results on the statistical analyses for the physiological measurements

	Heart rate	Skin conductance[6]
PRCS	$F(1,17) = 2.51$, $p = .132$	$F(1,15) = 0.57$, $p = .461$
Listening	$F(1,49) = 1.44$, $p = .236$	$F(1,39) = 4.59$, $p = .039$
Speaking	$F(1,49) = 0.18$, $p = .671$	$F(1,40) = 0.23$, $p = .638$
PRCS×Listening	$F(1,47) = 5.90$, $p = .019$	$F(1,38) = 0.26$, $p = .612$
PRCS×Speaking	$F(1,49) = 0.49$, $p = .487$	$F(1,34) = 0.37$, $p = .548$
Listening×Speaking	$F(1,49) = 0.04$, $p = .849$	$F(1,40) = 3.92$, $p = .055$
PRCS×Listening×Speaking	$F(1,49) = 0.49$, $p = .487$	$F(1,38) = 0.99$, $p = .325$

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 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).

[6] The effect of listening on skin conductance was not significant ($F(1,59) = 2.96$, $p = 0.091$) when the analysis was conducted using PRCS between-subjects variable with three levels.

Table 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 = .239
SAM-dominance	0 (1.0)	z = 1.48, p = .139
Dialogue length	21.0 (125.0)	z = 1.00, p = .317
Heart rate	-0.54 (4.28)	z = 0.14, p = .886
Skin conductance	9.7 (70.20)	z = 0.94, p = .346

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 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.

Table 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)	$z = 0.94$, $p = .345$
SAM-dominance	5.0 (3.0)	4.0 (4.0)	$z = 2.56$, $p = .011$
Dialogue length	298.5 (163.3)	251.8(188.1)	$z = 1.03$, $p = .304$
Heart rate	70.9 (13.7)	71.1 (13.3)	$z = 0.14$, $p = .989$
Skin conductance	205.4 (293.9)	203.7 (270.5)	$z = 1.10$, $p = .274$

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 (D. 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, Crane, & Fredrickson, 2010; Kleinsmith & 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.

Acknowledgements

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

Reference

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

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

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

Baylor, A. L., Ryu, J., & Shen, E. (2003). *The effects of pedagogical agent voice and animation on learning, motivation and perceived persona.* Paper presented at the Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications.

Beidel, D., Turner, S., Jacob, R., & Cooley, M. (1989). Assessment of social phobia: Reliability of an impromptu speech task. *Journal of anxiety disorders, 3*(3), 149-158.

Bradley, M. M., & Lang, P. J. (2007). The International Affective Digitized Sounds (IADS-2): Affective ratings of sounds and instruction manual. *Emotion*, 29-46.

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

Brinkman, W.-P., Hattangadi, N., Meziane, Z., & Pul, P. (2011). *Design and Evaluation of a Virtual Environment for the Treatment of Anger.* Paper presented at the Proceedings of Virtual Reality International Conference (VRIC 2011), Laval, France.

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

Brinkman, W.-P., van der Mast, C. A. P. G., Sandino, G., Gunawan, L. T., & 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., & Meyer, J.-J. (2012). *Virtual reality negotiation training increases negotiation knowledge and skill.* Paper presented at the Intelligent Virtual Agents.

Broekens, J., Qu, C., & Brinkman, W.-P. (2012). Dynamic Facial Expression of Emotion Made Easy. *Technical report* (pp. 1-30): Interactive Intelligence, Delft University of Technology.

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

Clark, D. A., & Beck, A. T. (2011). *Cognitive Therapy of Anxiety Disorders: Science and Practice*.

Core, M., Traum, D., Lane, H., & Swartout, W. (2006). Teaching negotiation skills through practice and reflection with virtual humans. *Simulation, 82*.

Cowell, A. J., & Stanney, K. M. (2003). *Embodiment and Interaction Guidelines for Designing Credible, Trustworthy Embodied Conversational Agents.* Paper presented at the 4th International Workshop on Intelligent Virtual Agents IVA 2003.

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., & Gratch, J. (2012). *The Effect of Virtual Agents' Emotion Displays and Appraisals on People's Decision Making in Negotiation.* Paper presented at the Intelligent Virtual Agents.

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

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

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

Gratch, J., & Okhmatovskaia, A. (2006). *Virtual rapport.* Paper presented at the Intelligent Virtual Agents.

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

Gregg, L., & 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., & 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., & Krijn, M. (2004). *Usability of therapist's user interface in virtual reality exposure therapy for fear of flying.* Paper presented at the In Proceedings of the Euromedia 2004.

Haring, M., Bee, N., & Andre, E. (2011). *Creation and Evaluation of emotion expression with body movement, sound and eye color for humanoid robots.* Paper presented at the RO-MAN, 2011 IEEE.

Irtel, H. (2007). PXLab: The Psychological Experiments Laboratory [online]. *Mannheim (Germany): University of Mannheim.*

James, L. K., Lin, C.-Y., Steed, A., Swapp, D., & Slater, M. (2003). Social anxiety in virtual environments: results of a pilot study. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society, 6*(3), 237-243.

Kleinsmith, A., & 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., & 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., & Nikolaidou, M. M. (2009). *Using affective avatars and rich multimedia content for education of children with autism.* Paper presented at the Proceedings of the 2nd International Conference on PErvsive Technologies Related to Assistive Environments - PETRA '09, New York, New York, USA.

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

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

Kulms, P., Kramer, N. C., Gratch, J., & Kang, S.-H. (2011). *It's in Their Eyes: A Study on Female and Male Virtual Humans' Gaze.* Paper presented at the Intelligent Virtual Agents.

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

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

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*.

Lanyi, C. S., Stark, J. A., Kamson, M. E., & 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., & 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., & 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).

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

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

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., & 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., & 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., & 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. *Standford, CA: Stanford University Press*.

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

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

Powers, M. B., & 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., & Smits, J. A. J. (2013). Do conversations with virtual avatars increase feelings of social anxiety? *Journal of anxiety disorders*. doi: 10.1016/j.janxdis.2013.03.003

Qiu, L., & 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., & Nass, C. (1996). *The Media Equation*.

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

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.*

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

Sinaceur, M., Adam, H., Van Kleef, G. a., & 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., & Clark, D. M. (2006). An experimental study on fear of public speaking using a virtual environment. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society, 9*(5), 627-633.

Swartout, W. R. (2006). *Simulators for human-oriented training.* Paper presented at the Proceedings - Winter Simulation Conference.

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

ter Heijden, N., Qu, C., Wiggers, P., & Brinkman, W.-P. (2010). *Developing a Dialogue Editor to Script Interaction between Virtual Characters and Social Phobic Patient.* Paper presented at the Proceedings of the ECCE2010 workshop - Cognitive engineering for technology in mental health care and rehabilitation.

Van Kleef, G. a., & Dreu, C. K. W. D. (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., & 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., & Higgins, J. J. (2011). The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only A NOVA Procedures. *CHI 2011 Session: Research Methods*, 143-146.

Wong, J. W.-E., & McGee, K. (2012). *Frown More, Talk More: Effects of Facial Expressions in Establishing Conversational Rapport with Virtual Agents.* Paper presented at the Intelligent Virtual Agents.

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