Public speaking in virtual reality
Audience design and speaker experiences

Kang, Ni

DOI
10.4233/uuid:e920dec8-2b71-4377-bc0f-8f9c950fff42

Publication date
2016

Document Version
Final published version

Citation (APA)
Kang, N. (2016). *Public speaking in virtual reality: Audience design and speaker experiences*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:e920dec8-2b71-4377-bc0f-8f9c950fff42

Public speaking in virtual reality

Audience design and speaker experiences

PUBLIC SPEAKING IN VIRTUAL REALITY

AUDIENCE DESIGN AND SPEAKER EXPERIENCES

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op woensdag 2 November 2016 om 12:30 uur

door

Ni KANG

Master of Science in Instrument Science and Technology from Southeast University, China,
Bachelor in Measuring Control Technology and Instruments from Southeast University, China,
geboren te Liaoning, China.

Dit proefschrift is goedgekeurd door de

promotor: prof. dr. M. A. Neerincx
copromotor: dr. W. P. Brinkman
copromotor: dr. M. B. van Riemsdijk

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. M. A. Neerincx,	Technische Universiteit Delft
Dr. W. P. Brinkman,	Technische Universiteit Delft
Dr. M. B. van Riemsdijk,	Technische Universiteit Delft
Onafhankelijke leden:	
Prof. dr. C. M. Jonker	Technische Universiteit Delft
Prof. dr. E. Eisemann	Technische Universiteit Delft
Prof. dr. J. J. Meyer	Utrecht Universiteit
Prof. dr. W. A. IJsselsteijn	Technische Universiteit Eindhoven
Prof. dr. A. Hanjalic	Technische Universiteit Delft, reservelid

The research reported in this thesis is supported by Chinese Scholarship Council (No. 2010609042).

ISBN 978-94-6186-745-2

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

For my beloved family.

Contents

Summary

Whether we are talking about our research at a conference, making a speech at a friend's wedding, or presenting a proposal in a business meeting, we have to speak in public from time to time. How well we deliver a presentation affects the way people think about us and our message. To deliver a well-received speech, preparation is necessary. Among various speech preparation activities, practicing with an audience is regarded as an effective way for enhancing speech performance. However, it is often impractical to organize an audience to practice a presentation and to arrange the diverse set of audience behaviours that are tailored to trainee's individual skills and learning goals. Virtual reality can provide a solution by practicing with a virtual audience. Although virtual audiences have been used in many domains, e.g., evoking social stress, therapy for social phobia, and improving teaching performance, little research has been reported on the impact of virtual audiences on public speakers' belief and performance. Therefore, this thesis aims to create a virtual audience which generates flexible expressive behaviours for a public speaking scenario and examines how public speaking experiences in front of such an audience affect the speakers' belief and speech performance.

To create an expressive virtual audience, the audience's body language should convey their affective and characteristic connotations, e.g., moods, attitudes, and personalities. To generate expressive audience behaviour adjustable at runtime, a parameterized audience model is proposed so that the audience's behaviour is controlled by model parameters that define virtual humans' moods, attitudes, and personalities. Due to the lack of a common, comprehensive and univocal knowledge-base on audiences' expressive behaviour, the parameterized behavioural model was created based on statistical models abstracted from observations of real audiences. Perception studies of the virtual audience were conducted. The results showed that people could recognize different attitudes expressed by the virtual audience, and they could also perceive the changes in some of the mood, personality, and attitude parameters of the virtual audience. Thus, the parameterized audience model could generate expressive behaviour of a virtual audience that can be recognized accurately by humans.

To design an expressive virtual audience used for a certain purpose, e.g., scientific research or skills training, developers need to know how specific audience behaviours are associated with certain characteristics of an audience, such as attitude, and how well people can recognize these characteristics. To contribute to such knowledge, this thesis explored the potential of constructing audience variations using the parameters of the behavioural model and people's understanding on audience behavioural styles. People were firstly invited to design various audience scenarios and, subsequently, the design was validated in a perception study. The results showed that individuals had consistent conception on audience behaviour and recognized audience styles for a variety of audience scenarios. To gain further insight about people's perception of the audience styles, the audience behaviour was investigated for different audience types. By examining the statistical models for generating audience behaviour, specific behaviours for different audience styles were identified, e.g., an attentive audience tended to look forward, critical audience often showed a more closed posture, and a bored, impatient audience fidgeted in their seats. . These findings provided insight on the behavioural design of virtual audiences.

To investigate whether public speaking experiences obtained in virtual environments could affect speakers' beliefs and performance, two empirical studies were conducted respectively on direct experiences and indirect or vicarious experiences of public speaking in virtual reality. While practicing with a real audience is suggested to be effective in enhancing speech performance, such practice is not commonly followed due to many practical problems. Those problems can be solved by the direct experience of public speaking in virtual reality, i.e., speaking in front of a virtual audience. This thesis sets out to study this practicing technique for enhancing public speaking performance, focusing especially on its effectiveness and people's training satisfaction. This practicing technique was investigated in a course for public speaking skills training and was compared with speaking in front of an imaginary audience. The results showed that individuals seemed to benefit more from a virtual audience than an imaginary audience in reducing speech anxiety. They were also more positive towards training with a virtual audience regarding both the training process and its effect on their presentation ability.

This thesis also explored an indirect experience in virtual reality. The indirect experience, which mixed the features of both direct experience and traditional vicarious experience, enabled individuals to observe one's experience from a first person perspective without performing by themselves. Although direct experiences are considered more influential than vicarious experiences in affecting individuals, an individual who holds a weak efficacy belief may have difficulty to experience directly. As the experience is from a first person perspective, such an experience may help individuals enhance self-efficacy better than a traditional vicarious experience. In another aspect, such experiences may share similarities with traditional vicarious experiences, regarding the moderating effect of the observed model or performer's identity during the experience. Therefore, this thesis investigated the effect of the first-person-perspective vicarious experience and the moderating effect of self-identification in such experiences. An empirical study was carried out by comparing experiences of public speaking with a high level of self-identification and experiences with a low level of self-identification. The findings indicate that the first-person-perspective experiences with a higher level of self-identification are more effective in modifying individuals' self-efficacy beliefs in public speaking than experiences with a lower level of self-identification. Additionally, the moderating effect of self-identification was found on the relationship between self-efficacy after the vicarious experience and the model's performance in the experience. The moderating effect also existed on the association between the vicarious experience and a direct experience. Only when a high level of self-identification was present, was the vicarious experience similar to a direct experience with regard to the experienced presence and their effects on self-efficacy scores.

In short, this thesis demonstrates the possibility of generating virtual audiences with adjustable and recognizable expressive behavioural styles. The experience, either direct or indirect, of presenting in front of a virtual audience can affect the presenters' efficacy beliefs. Moreover, practicing in front of the virtual audience helps to reduce speech anxiety, and the satisfaction this method provides may motivate individuals to practice their presentation. The effects are important, as well-delivered presentations can have personal benefits, or, in the case of the iconic speeches, change the course of history.

SAMENVATTING

Of we nu ons onderzoek presenteren bij een conferentie, een speech geven bij het huwelijk van een vriend of een voorstel toelichten in een zakelijke vergadering, velen van ons moeten nu en dan in het openbaar spreken. Hoe goed we een presentatie brengen heeft invloed op de manier waarop mensen over ons en onze boodschap denken. Om een goede speech te kunnen geven is voorbereiding nodig. Van de verschillende manieren om je voor te bereiden wordt oefenen met publiek gezien als een effectieve manier om een betere presentatie te geven. Het is echter vaak onpraktisch om publiek te regelen om je presentatie te oefenen en ervoor te zorgen dat het publiek zich zo gedraagt dat het past bij wat je kunt en wat je wilt oefenen.

Virtual reality kan een oplossing bieden hiervoor doordat het de mogelijkheid biedt te oefenen met een virtueel publiek. Virtuele karakters worden al in veel domeinen in virtual reality ingezet, zoals voor het opwekken van sociale stress, voor de behandeling van sociale fobie en voor het verbeteren van vaardigheden in lesgeven. Echter, er is nog weinig onderzoek gedaan naar de impact van een virtueel publiek op de presentatievaardigheden en het zelfeffectiviteitsbeeld van iemand met betrekking tot het geven van een presentatie. Met "zelfeffectiviteitsbeeld"bedoelen we het vertrouwen dat iemand heeft in het eigen kunnen met betrekking tot het uitvoeren van een specifieke taak, in dit geval spreken in het openbaar. Dit proefschrift stelt als doel het creëren van een virtueel publiek dat flexibel expressief gedrag kan vertonen om spreken in het openbaar met verschillende soorten publiek te oefenen. Daarnaast wordt onderzocht hoe spreken voor een dergelijk virtueel publiek het zelfeffectiviteitsbeeld en de presentatievaardigheden beïnvloedt.

Om een expressief virtueel publiek te creëren moet de lichaamstaal van de virtuele karakters in het publiek verschillende soorten emoties, attitudes (zoals een kritische houding) en persoonlijkheden kunnen laten zien. Om dit soort gedrag automatisch te kunnen genereren en varieren tijdens het oefenen in de virtuele omgeving, hebben we een gedragsmodel ontwikkeld met verschillende parameters. Het gedrag van het publiek wordt gecontroleerd door parameters die emoties, attitudes en persoonlijkheid definiëren. Omdat er geen algemeen beschikbare volledige set van gegevens met expressief publieksgedrag was, hebben we observaties uitgevoerd van “echt” publiek en een gedragsmodel gecreëerd op basis van statistische modellen van de observatiegegevens. Vervolgens hebben we perceptiestudies uitgevoerd om te onderzoeken hoe mensen dit expressieve virtuele publiek waarnemen. De resultaten laten zien dat mensen de verschillende attitudes van het virtuele publiek konden herkennen, evenals een aantal veranderingen in de emotie-, persoonlijkheids- en attitude-parameters. Oftewel, het parametrische gedragsmodel kan expressief gedrag genereren voor een virtueel publiek dat correct herkend kan worden door mensen.

Om een expressief virtueel publiek te ontwerpen dat gebruikt kan worden voor een specifiek doel, zoals wetenschappelijk onderzoek of training, moeten we weten hoe spe-

cifiek publieksgedrag geassocieerd is met eigenschappen van een publiek, zoals attitude, en hoe goed mensen deze verschillende soorten publiek kunnen herkennen. Om dit te onderzoeken hebben we bestudeerd of het mogelijk is om verschillende publieksstijlen te simuleren met behulp van het gedragsmodel, en in hoeverre mensen deze stijlen herkennen. Dit onderzoek is uitgevoerd in verschillende stappen. Allereerst hebben we mensen gevraagd om verscheidene scenario's van spreken in het openbaar te ontwerpen. Vervolgens is het ontwerp van deze scenario's gevalideerd in een perceptiestudie. De resultaten laten zien dat mensen consistent dachten over het publieksgedrag en dat ze de publieksstijlen herkenden in de verschillende scenario's. Om meer inzicht te krijgen in de perceptie van publieksstijlen hebben we het (uiterlijke) gedrag van het virtuele publiek onderzocht voor de verschillende publieksstijlen. Door het gedragsmodel te bestuderen werden specifieke gedragingen voor verschillende publieksstijlen geïdentificeerd. Zo kijkt een aandachtig publiek meer naar voren, laat een kritisch publiek vaak een gesloten houding zien, en wiebelt een verveeld ongeduldig publiek vaker met de handen en benen. Deze resultaten geven inzicht in de mogelijkheden die de technologie biedt in het vormgeven van gedrag van virtueel publiek.

Om te onderzoeken of de ervaring met het spreken voor virtueel publiek invloed kan hebben op het zelfeffectiviteitsbeeld en de prestatie in het geven van een presentatie zijn twee empirische studies uitgevoerd. Hierbij is een onderscheid gemaakt tussen de directe ervaring met het actief zelf spreken voor een publiek, en de indirecte passieve ervaring met een door computer gesimuleerde beleving van het spreken voor een publiek. Dit proefschrift bestudeert deze directe oefentechniek voor het verbeteren van spreken in het openbaar en focust specifiek op de effectiviteit en tevredenheid van mensen met de training. Deze oefentechniek is onderzocht in een cursus voor openbaar spreken en vergeleken met spreken voor een denkbeeldig publiek waarbij sprekers zich in gedachten voorstellen dat ze voor een publiek staan. De resultaten laten zien dat mensen meer baat hadden van een virtueel publiek dan een denkbeeldig publiek als het gaat om het verlagen van spreekangst. Ze waren ook positiever over het trainen met een virtueel publiek wat betreft het trainingsproces en het effect op hun presentatiecapaciteiten.

Een probleem met het opdoen van een directe ervaring van spreken voor publiek is dat mensen een laag zelfeffectiviteitsbeeld hier erg tegenop kunnen zien. Daarom wordt in dit proefschrift ook de indirecte ervaring van het gesimuleerd spreken in een virtuele omgeving onderzocht. Deze indirecte ervaring heeft eigenschappen van zowel een directe ervaring als een plaatsvervangende ervaring, een ervaring die mensen opdoen door bijvoorbeeld te kijken hoe andere personen het doen. De indirecte ervaring stelt personen in staat om de eigen ervaring te observeren vanuit een eerste-persoons perspectief zonder zelf iets te doen. Deze ervaring is gecreeerd door de proefpersoon in de virtuele omgeving in de huid van een virtueel karakter te laten kruipen, waarbij de persoon beleeft hoe dit karakter een presentatie geeft voor een virtueel publiek. De mate waarmee een persoon zich identificeerde met dit karakter kan een modererende invloed hebben op een dergelijke ervaring. Om deze reden bestudeert dit proefschrift het effect van de indirecte ervaring in het eerste-persoons perspectief en de modererende invloed van zelfidentificatie bij zulke ervaringen. Een empirische studie is gedaan door de ervaring van spreken in het openbaar te vergelijken bij een hoog en een laag niveau van zelfidentificatie. De resultaten wijzen erop dat het gevoel van bekwaamheid meer beinvloed werd als de zelfidentificatie hoger was. Bovendien werd de modererende invloed van zelfidentificatie gevonden op de relatie tussen het zelfeffectiviteitsbeeld na de indirecte ervaring en de perceptie over hoe goed de presentatie uitgevoerd was. Enkel wanneer een hoog niveau van zelfidentificatie aanwezig was, was de indirecte ervaring vergelijkbaar met een directe ervaring wat betreft de ervaren aanwezigheid in de virtuele wereld en de invloed op het zelfeffectiviteitsbeeld.

Samengevat demonstreert dit proefschrift de mogelijkheid van het genereren van virtueel publiek met aanpasbare en herkenbare expressieve gedragsstijlen. De ervaring, direct of indirect, van het presenteren voor een virtueel publiek kan het zelfeffectiviteitsbeeld van de presentator beïnvloeden. Bovendien helpt het oefenen voor een virtueel publiek om spreekangst te verminderen. Bovendien oefenen mensen liever met een virtueel publiek dan met een denkbeeldig publiek. Gebruik maken van een virtueel publiek kan mensen dus motiveren om meer te oefenen, en oefenen verbetert de prestatie. Deze effecten zijn belangrijk omdat een goede presentatie persoonlijke voordelen kan hebben of, in het geval van iconische speeches, de loop van de geschiedenis kunnen veranderen.

1

INTRODUCTION

Four score and seven years ago our fathers brought forth on this continent, a new nation, conceived in liberty, and dedicated to the proposition that all men are created equal.

– "The Gettysburg Address" by Abraham Lincoln, 19 November, 1863

You ask, what is our aim? I can answer in one word. It is victory. Victory at all costs – Victory in spite of all terrors – Victory, however long and hard the road may be, for without victory there is no survival.

- "Their Finest Hour" by Winston Churchill, 18 June, 1940

I have a dream that one day down in Alabama, with its vicious racists, with its governor having his lips dripping with the words of interposition and nullification – one day right there in Alabama little black boys and black girls will be able to join hands with little white boys and white girls as sisters and brothers.

- "I Have a Dream" by Martin Luther King Jr., 28 August, 1963

1.1. MOTIVATION

GREAT speeches change our world. The Gettysburg Address by Abraham Lincoln changed the way people thought about the Constitution and has become an authoritative expression of the American spirit. These 273 words have remade America. When confronted with the threat of invasion from Nazi-occupied France, Winston Churchill rallied the nation for the Battle of Britain to come and showed their resolve to fight. The hope, courage, and determination from the speech saved Great Britain. Martin Luther King's speech advanced the civil rights legislation of America and brought his messages worldwide. The messages set off a worldwide movement for equal rights.

Although those famous public addresses are iconic and rare, almost everyone has to speak in public from time to time, e.g., talking about research at a conference, making a speech at a friend's wedding, or presenting a proposal in a team meeting. Although not necessarily as influential as those greatest speeches, how well we present affects the way people think about us. For example, people tend to regard presenters as more credible and intellectual when they have more eye contact with their audience during presentations [1], [2], and people are more likely to believe the presenters and be persuaded when the speech is fluent and well organized [3]. People's attitudes change more in the direction of persuasive messages when supported by strong arguments instead of weak arguments [4].

To deliver a well-received speech, a good preparation is first of all needed. Among all the speech preparation activities, Menzel and Carrell [5] specifically found that more rehearsals in front of an audience and less anxiety are associated to better speech performance. Moreover, the study by Ayres et al. [6] indicated that practicing with an audience helps to decrease public speaking anxiety and increase the willingness to speak. Smith and Frymier [7] further found that students who had practiced with an audience performed better than those who had practiced without an audience. Hence, practicing in front of an audience can be an effective method to improve the performance.

However, it is often impractical to organize an audience to practice a presentation. Moreover, some people with high levels of speech anxiety may be reluctant or even unable to present in front of an audience. Virtual reality may provide a solution, e.g., practicing with a virtual audience, which are made up of a group of virtual humans. Besides the logistic advantage of not needing to arrange audience members and a suitable location, virtual reality also offers a unique ability to control the members of the virtual audience. For example, to study the effects of a virtual audience on anxiety, virtual audiences have been manipulated to be supportive or unsupportive [8]. In contrast, it is often impossible for real humans to replicate exactly a certain scenario. Because of these benefits, exposure in virtual reality is also being used as part of a treatment for anxiety disorders. Here patients are exposed to situations they fear. Exposure in virtual reality was found more acceptable for the patients than exposure in real situations [9]. Furthermore, virtual reality exposure has also been found to be an effective treatment method for anxiety disorders [10–13]. Therefore, this thesis aims to explore the design and creation of a virtual audience which can be manipulated to meet the momentary individual needs of patients or trainees.

Many applications may benefit from an expressive virtual audience. For example, treatment manuals of exposure therapy for social anxiety [14], [15] suggest controlling the audience attitude as an effective means of controlling anxiety in a public speaking scenario; studies on stress responses explored variations of stress tests using supportive and non-supportive audiences [8], [16]; the virtual audience in a public speaking training system manifested different attitudes as feedback for the speech performance [17]. Currently expressive virtual audiences are often represented by 3D models animated by a predefined script [18], or videos of actual people embedded in Virtual Environments (VEs) [19]. To control the audience's behaviour, different animations or videos should be prepared so that operators can switch between these clips. However, preparations may require considerable effort and thus are usually made in advance because explicit behaviours need to be scripted along the timeline for each audience member. Due to the effort involved, these pre-scripted animations and videos are often relatively short, causing the virtual audience to behave in repeating loops. This repetition may reduce the behavioural realism, thereby lowering the desired effect, e.g., lowering treatment efficacy [20]. Therefore, from an engineering perspective, a more flexible and efficient system needs to be developed for the virtual audience to generate expressive behaviours automatically. In this way, the audience's behavioural styles can be adapted in real time to meet the users' needs.

Once an expressive virtual audience is created, the question how the virtual audience can be used to affect humans becomes essential. Like real human audiences, virtual audiences have the ability to elicit responses in humans. When individuals are exposed to a virtual environment and perform in front of a virtual audience, their anxiety and performance can be affected. For example, Wallergard et al. [21] suggested that virtual audiences, as part of a stress test, can indeed, like human audiences, induce stress. Much research [22–24] has already focused on giving people the experience of performing in front of an audience as part of exposure therapy for individuals with social anxiety disorder. This experience has also benefited non-clinical applications. For example, Bautista and Boone [25] let teachers be trained with virtual students to master their skills of content delivery and student management. Likewise, Bissonnette et al. [26] trained performance arts students, in this case, young musicians, to overcome their performance anxiety by performing in front of a virtual audience. Therefore, because of its generic application, this thesis focuses specifically on the context of public speaking and investigates how experiences with a virtual audience affect the presenter.

1.2. Main Research Question and Hypothesis

The following main question has driven the research presented in this thesis:

> *Within public speaking scenarios, how can an expressive virtual audience be created and how do the experiences with a virtual audience affect the presenter?*

As illustrated in Figure 1.1, to study this question it required the implementation of an expressive virtual audience. Only then, it is possible to address the second part of the research question as how people perceive such expressive virtual audience, i.e., what attitude, mood, personality etc a virtual audience can convey and whether people can recognize variations in these audience characteristics. It then becomes possible to focus on the presenters' experience, their satisfaction of presenting for such an audience,

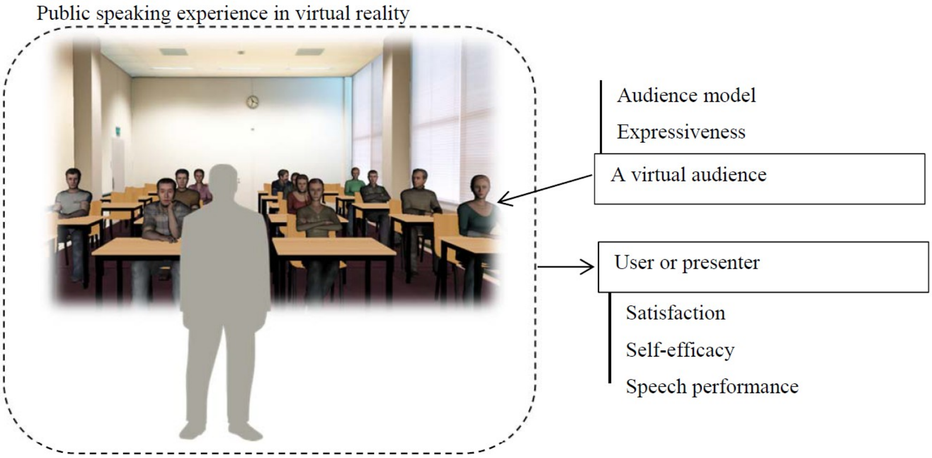


Figure 1.1: Illustration of the topic of this thesis

but also how it impacts their performance, and their belief about their ability to give a talk, i.e. their self-efficacy. Thus, the first objective of this thesis is to develop a parameterized audience model that can generate and adapt the expressive behaviours of virtual audiences in real time. Support for this idea can be found in work of Wang et al. [27]. They showed that a model of virtual listeners' feedback behaviour in a multiparty conversation was able to take into account several factors which affect the listeners' behaviour, such as their conversational roles and goals, understanding, and attitudes. Furthermore, Busso et al. [28] demonstrated the possibility of computational models to predict a speaker's head motion for different emotions, and their evaluations suggested that these models successfully emphasized the emotional content and improved the virtual speaker's behavioural naturalness. These examples show the potential of a parameterized model to generate expressive behaviours for a virtual audience. Additionally, as suggested by Allwood [29], human's nonverbal behaviour exhibits clues about a person's mental and physiological states, such as moods, attitudes, and personality. Thus, this thesis proposes a parameterized agent model for the audience to generate expressive listening behaviour, i.e. behavioural styles. These styles can be modulated by adjusting the agent's attributes such as mood, attitude, personality, and energy level. Virtual characters' behaviour can be generated autonomously mainly in two ways: (1) computed by crafted rules that specify which behaviour should be generated in a certain context based on psychological knowledge and literature [30], or (2) generated by statistical models that predict body postures based on observation [31]. Whereas the statistical approach needs real-life observations of a certain phenomenon to build a model, the theoretical approach requires more broad and general knowledge of the phenomena. At this moment, complete, coherent, and formal specifications of audience behaviour cannot be derived from theories; hence, the statistical approach was applied in this thesis to generate the virtual audience's behaviour.

An important objective for the generation of expressive virtual audiences is that people can notice variations in the generated behavioural styles, and furthermore, can recognize distinct styles. When it comes to people's ability to recognize the mental and physiological states simulated by a single virtual human, considerable work has been done. For example, individuals can recognize affective states from simulated facial expressions [32]. Besides emotions, Chollet et al. [31] and Hu, Walker, Neff, and Tree [33] demonstrated that people can recognize a virtual human's attitude and even personality expressed by full body postures. Besides work that focuses on recognition based on the behaviour of a single virtual human, some work also focuses on groups of virtual humans. For example, Prada and Paiva [34] developed a model supporting group dynamics for autonomous agents to perform in groups with human users. Group behaviour of virtual agents has also been modeled based on the agents' interpersonal relationship [35]. Ravenet et al. [36] modeled nonverbal behaviour of virtual agents in a conversational group and demonstrated that individuals recognized the expressed interpersonal attitudes of the agents. Some research specifically focuses on behaviour of virtual audiences in a public speaking scenario. Poeschl and Doering [37] and Tudor et al. [38] provided initial guidelines for behavioural design of realistic virtual audiences. They observed the behaviour of a typical audience in a lecture and explored the behavioural patterns such as frequency, duration, and postural sequence of certain behaviour categories, e.g. paying attention.

In addition to the perception and recognition of group attitude and an overall mental state, a virtual audience can have an effect on presenters' experience when delivering their talks. The effect can manifest itself in several ways. Most noticeable and fundamental is the reported feeling of being present in front of a virtual human or a virtual audience. Studies have shown that people can experience a sense of presence when interacting with a virtual human [39], exposed to a group of virtual classmates [40], and speaking in front of a virtual audience [41]. Another effect presenters can experience is a heightened level of anxiety, stress, or arousal. For example, people have reported anxiety when speaking to a virtual character [39], [42], [43], speaking among a group of virtual characters [40], or giving a presentation in front of a virtual audience [44]. The presence of a virtual audience has also been found to affect people as a real audience did. For example, individuals were found inhibited when performing a complex task with the presence of either a real or virtual audience [45]. When musicians performed in front of a virtual audience, their anxiety responses were similar to those elicited by a real audience. Other than an emotion impact, exposure to virtual humans has also been found to affect beliefs such as self-efficacy [40]. Besides the immediate effect during exposure, exposure might also have long lasting impact. This is of course essential for educational and therapeutic use. Morina et al. [24] and Anderson et al. [22] found that the treatment gains were maintained months after exposure.

It should be noted that, without a visible audience, an imaginary audience can also affect people's emotion and performance. For example, imaginary audiences and virtual audiences were found similarly effective in evoking social anxiety [16]. Both methods may help to reduce people's anxiety during presentation and improve public speaking performance. However, practicing with an imaginary audience holds a number of drawbacks when comparing practicing with a virtual audience, e.g., the limitations of an imagery task such as requiring considerable attentional resources and lack of control [46], which potentially makes it less satisfying, efficient and effective for some individuals.

For some phobic individuals, it might initially still be very challenging to practice directly in front of a virtual audience. Passively observing initially someone else speaking would however be more attainable. This thesis argues that even observing a public speaking experience from the presenter's perspective in virtual reality, can still affect the observer. Several studies have shown that indirect or vicarious experiences obtained in virtual reality may affect one's belief and behaviour. For example, a person's self-efficacy can be weakened when he is observing virtual classmates praising other virtual classmates but negatively criticizing him when he answers questions in a virtual classroom [40]. People were also found doing more physical exercises after observing a jogging virtual lookalike, i.e., a virtual-self, than observing a dissimilar virtual character jogging [47]. The work in this thesis goes one step further, by creating a first-person-perspective vicarious experience. This experience mixes the features of direct experiences and traditional vicarious experiences. As individuals experience a scenario from a first person perspective without actively performing the behaviour in question, the vicarious experience gives them the experience of accomplishment and thereby influencing their future performance. On the other hand, this experience is in some aspects also similar to a traditional vicarious experience. A person only experiences passively a scenario that unfolds automatically, i.e., the person only observes and does not need to act during the experience. When learning from traditional vicarious experiences, besides modelled performance, people may also evaluate their own capability by comparing themselves to the model on personal characteristics such as age and gender, when they are assumed to be predictive of performance capabilities. For example, children have been reported to derive a stronger self-efficacy from peer modelling than observing adult models exemplifying the same task [48]. Thus, learning can be more effective, or the modelled performance is more relevant to a person, if the person perceives more similarity between oneself and the model [49]. Thus, the identification of the model of the first-person-perspective vicarious experience, i.e., how individuals relate the experience to themselves, can be a moderator on the effects of such an experience. In other words, a higher sense of self-identification during the experience can be more effective in modifying one's efficacy belief.

In conclusion, the thesis distinguishes the following four hypotheses:

H1. A parameterized audience model can generate a virtual audience with expressive behaviour.

H2. People can recognize different styles of the expressive behaviours generated by the audience model.

H3. People prefer practicing with a virtual audience over practicing with an imaginary audience.

H4. The level of self-identification affects self-efficacy after a first-person-perspective vicarious experience.

1.3. RESEARCH APPROACH

To test the first hypothesis, a parameterized agent model for virtual audiences was first created with the ability to generate expressive behaviours. The model provided control over the audience behaviour by setting parameters such as attitude, mood, and personality of virtual audience members. For this, real audiences were observed in four different presentations, designed to elicit various attitudes, moods, and physical states of the audiences. During the observation of audiences' behaviour, the audiences' states, e.g., attitude and mood, were collected by questionnaires. To establish the connection between the audiences' states and specific behaviours, statistical models were abstracted from the observed data, with the audience states as predictor parameters and audience behaviour as response variables. To investigate whether individuals could recognize the generated expressive behaviour, two perception studies were conducted. In the first experiment, 22 participants were exposed for 12 minutes to a virtual audience while every minute the model's parameter setting was systematically changed. Throughout the exposure, participants were asked to describe orally the states of the audience without restrictions. In the second experiment, another 22 participants were exposed to the virtual audience in the same sequence of 12 conditions. After watching the audience in each condition, participants were asked to rate the virtual audience by scoring parameter dimensions used by the model, e.g. attitude, and mood. The model and empirical study are described in Chapter 2.

Examination of the second hypothesis centred on the perception validation of the virtual audience expressiveness. Two questions were at the core of the validation: when changing the model parameters, are people able to notice variation in audience behaviour, and can people recognize distinct behavioural audience styles? Four studies were conducted. The first study aimed to investigate individuals' perception on audience characteristics, in particular, mood, personality, and attitude, which result in variations in audience behaviour. Hence, a paired comparison perception experiment was conducted. In the experiment, pairs of virtual audiences with different parameter settings were shown to participants, who were required to judge a specific quantitative difference in an audience property, e.g. higher or lower arousal. In the following two studies, people's understanding on audience behavioural styles was explored. Thus, two experiments respectively explored the design and the perception of various audience scenarios. In the first experiment, people were invited to use the noticeable audience parameters earlier identified to design audience behaviour for a set of public speaking scenarios, such as giving a best man's speech at a wedding, or presenting a business proposal to a number of potential investors. Based on the parameter settings made, audience scenarios were clustered on their similarity into general behavioural audience styles. To validate that people indeed recognise these general behavioural audience styles, the second experiment invited people to match videos of virtual audiences exhibiting these styles with descriptions of these behavioural audience style. The last step in perception validation was to go back to the original underlying model and determine what specific audience postures and behaviours made up these general behavioural audience styles. The four studies and corresponding results are presented in detail in Chapter 3.

Using the described virtual audience with its recognizable behavioural styles, various public speaking scenarios could be constructed, which provided opportunities for individuals to obtain a variation of public speaking experiences. It also provided a basis for testing the third hypothesis, regarding the effect a virtual audience has on presenters. A public speaking training course was organized on the university campus, which included three training sessions followed by a closing presentation where participants gave their presentation to a human audience. To acquire direct experiences as a public speaker, participants were asked to practice their own presentations in front of a virtual audience in three training sessions. This experiment employed a between-subjects design, which meant that another participant group received the same training content but were required to practice with an imaginary audience in the training sessions. In all sessions, participants' heart rate, subjective anxiety level, and self-efficacy in public speaking were collected. In the final presentation, two human audience members also scored the participants' speech performance. Furthermore, participants scored their satisfaction with the practicing techniques. The results of the experiment gave an insight into the practicing technique's effectiveness and people's satisfaction. A detailed description of the study is presented in Chapter 4.

The fourth hypothesis was also tested in an empirical study. Vicarious experiences were constructed from a public speaker's perspective where the speaker was giving lectures on elementary arithmetic in front of a virtual audience. To exploit the moderating effect of self-identification on the experience, experiences with a high level of self-identification were compared with experiences with a low level of self-identification in a between-subjects experiment (n = 60). After being exposed to the vicarious experience, participants from both conditions actively gave another lecture on elementary arithmetic. Here the vicarious experiences' effect was compared between the two levels of self-identification. The investigation analysed individuals' experience, such as self-efficacy, presence response, and anxiety level, which were collected by questionnaires in both vicarious experience and direct experience of the experiment. To investigate the moderating effect of self-identification on such vicarious experiences, correlation was examined between self-efficacy after the vicarious experience and the performance of the model in the experience. Besides, the moderating effect of self-identification was also examined on the association between the vicarious experiences and direct lecture experiences regarding the experienced presence and the experiences' effect on self-efficacy. This study and its findings are presented in detail in Chapter 5.

The conclusions that can be drawn from the various studies presented in this thesis are discussed in chapter 6. The main contributions of the research and suggestions for future research are also put forward here.

REFERENCES

[1] P. Wu, K. Yoshida, H. Suto, S. Hanita, T. Yokoi, and N. Yamaji, "The Effect of Speaker' s Behaviors on Audiences' Impressions," in *The 5th International Congress of the International Association of Societies of Design Research: IASDR 2013*, 2013.

[2] S. A. Beebe, "Eye contact: A nonverbal determinant of speaker credibility," *The Speech Teacher*, vol. 23, no. 1, pp. 21–25, Jan. 1974.

[3] J. C. McCroskey and R. S. Mehrley, "The effects of disorganization and nonfluency on attitude change and source credibility," *Speech Monographs*, vol. 36, no. 1, pp. 13–21, Mar. 1969.

[4] H. S. Park, T. R. Levine, C. Y. Kingsley Westerman, T. Orfgen, and S. Foregger, "The

Effects of Argument Quality and Involvement Type on Attitude Formation and Attitude Change: A Test of Dual-Process and Social Judgment Predictions," *Human Communication Research*, vol. 33, no. 1, pp. 81–102, Jan. 2007.

[5] K. E. Menzel and L. J. Carrell, "The relationship between preparation and performance in public speaking," *Communication Education*, vol. 43, no. 1, pp. 17–26, Jan. 1994.

[6] J. Ayres, T. Schliesman, and D. A. Sonandré, "Practice makes perfect but does it help reduce communication apprehension?," *Communication Research Reports*, vol. 15, no. 2, pp. 170–179, Mar. 1998.

[7] T. Smith and A. B. Frymier, "Get 'Real': Does Practicing Speeches Before an Audience Improve Performance?," *Communication Quarterly*, vol. 54, no. 1, pp. 111–125, Feb. 2006.

[8] S. E. Taylor, T. E. Seeman, N. I. Eisenberger, T. a Kozanian, A. N. Moore, and W. G. Moons, "Effects of a supportive or an unsupportive audience on biological and psychological responses to stress.," *Journal of personality and social psychology*, vol. 98, no. 1, pp. 47–56, Jan. 2010.

[9] A. Garcia-Palacios, C. Botella, H. Hoffman, and S. Fabregat, "Comparing Acceptance and Refusal Rates of Virtual Reality Exposure vs. In Vivo Exposure by Patients with Specific Phobias," *CyberPsychology & Behavior*, vol. 10, no. 5, pp. 722–724, Oct. 2007.

[10] T. D. Parsons and A. A. Rizzo, "Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 39, no. 3, pp. 250–261, Sep. 2008.

[11] M. Powers and P. Emmelkamp, "Virtual reality exposure therapy for anxiety disorders: A meta-analysis," *Journal of Anxiety Disorders*, vol. 22, no. 3, pp. 561–569, 2008.

[12] D. Opriş, S. Pintea, A. García-Palacios, C. Botella, Ş. Szamoskozi, and D. David, "Virtual reality exposure therapy in anxiety disorders: a quantitative meta-analysis," *Depression and anxiety*, vol. 29, no. 2, pp. 85–93, Feb. 2012.

[13] N. Morina, H. Ijntema, K. Meyerbröker, and P. M. G. Emmelkamp, "Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments," *Behaviour Research and Therapy*, vol. 74, pp. 18–24, Nov. 2015.

[14] R. G. Heimberg and R. E. Becker, *Cognitive-Behavioral Group Therapy for Social Phobia: Basic Mechanisms and Clinical Strategies.* Guilford Publications, 2002, p. 334.

[15] S. F. Hofmann and M. W. Otto, *Cognitive Behavioral Therapy for Social Anxiety Disorder: Evidence-Based and Disorder-Specific Treatment Techniques*, 1st ed. Routledge, 2008.

1

[16] O. Kelly, K. Matheson, A. Martinez, Z. Merali, and H. Anisman, “Psychosocial stress evoked by a virtual audience: Relation to neuroendocrine activity,” *CyberPsychology & Behavior*, vol. 10, no. 5, pp. 655–62, Oct. 2007.

[17] M. Chollet, G. Sratou, and A. Shapiro, “An interactive virtual audience platform for public speaking training,” in *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, 2014, pp. 1657–1658.

[18] D. P. Pertaub, M. Slater, and C. Barker, “An experiment on public speaking anxiety in response to three different types of virtual audience,” *Presence: Teleoperators & Virtual Environments*, vol. 11, no. 1, pp. 68–78, 2002.

[19] M. J. Gallego, P. M. G. Emmelkamp, M. Van der Kooij, and H. Mees, “The effects of a Dutch version of an Internet- based treatment program for fear of public speaking[202F?]: A controlled study,” *Int J Clin Health Psychol*, vol. 11, no. 3, pp. 459–472, 2011.

[20] V. Vinayagamoorthy, A. Steed, and M. Slater, “Building Characters[202F?]: Lessons Drawn from Virtual Environments,” in *Proceedings of Toward Social Mechanisms of Android Science: A CogSci 2005 Workshop*, 2005, pp. 119–126.

[21] M. Wallergård, P. Jönsson, G. Johansson, and B. Karlson, “A Virtual Reality Version of the Trier Social Stress Test: A Pilot Study,” *Presence: Teleoperators and Virtual Environments*, vol. 20, no. 4, pp. 325–336, Aug. 2011.

[22] P. L. Anderson, M. Price, S. M. Edwards, M. a Obasaju, S. K. Schmertz, E. Zimand, and M. R. Calamaras, “Virtual reality exposure therapy for social anxiety disorder: a randomized controlled trial.,” *Journal of consulting and clinical psychology*, vol. 81, no. 5, pp. 751–60, Oct. 2013.

[23] D. Hartanto, W.-P. Brinkman, I. L. Kampmann, N. Morina, P. G. M. Emmelkamp, and M. A. Neerincx, “Design and Implementation of Home-Based Virtual Reality Exposure Therapy System with a Virtual eCoach,” in *Proceeding of the 15th International Conference on Intelligent Virtual Agents*, 2015, vol. 9238, pp. 287–291.

[24] N. Morina, W.-P. Brinkman, D. Hartanto, I.L. Kampmann, and P.M.G. Emmelkamp, “Social interactions in virtual reality exposure therapy: A proof-of-concept pilot study.,” *Technology and health care: Official Journal of the European Society for Engineering and Medicine*, vol. 23, no. 5, pp. 581–9, Sep. 2015.

[25] N. U. Bautista and W. J. Boone, “Exploring the Impact of TeachME^{TM} Lab Virtual Classroom Teaching Simulation on Early Childhood Education Majors’ Self-Efficacy Beliefs,” *Journal of Science Teacher Education*, vol. 26, no. 3, pp. 237–262, Feb. 2015.

[26] J. Bissonnette, F. Dubé, M. D. Provencher, and M. T. Moreno Sala, “Virtual Reality Exposure Training for Musicians: Its Effect on Performance Anxiety and Quality.,” *Medical problems of performing artists*, vol. 30, no. 3, pp. 169–77, Sep. 2015.

[27] Z. Wang, J. Lee, and S. Marsella, "Towards more comprehensive listening behavior: Beyond the bobble head," in *Proceedings of the 11th international conference on Intelligent Virtual Agents*, 2011, vol. 6895, pp. 216–227.

[28] C. Busso, Z. Deng, and M. Grimm, "Rigid head motion in expressive speech animation: Analysis and synthesis," *IEEE Tansactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1075–1086, 2007.

[29] J. Allwood, "Bodily communication dimensions of expression and content," in *Multimodality in Language and Speech Systems*, B. Granstrom, D. House, and I. Karlsson, Eds. Kluwer Academic Publishers, 2002, pp. 7–26.

[30] E. Bevacqua, E. Sevin, S. J. Hyniewska, and C. Pelachaud, "A listener model: Introducing personality traits," *Journal on Multimodal User Interfaces*, vol. 6, no. 1–2, pp. 27–38, Apr. 2012.

[31] M. Chollet, M. Ochs, and C. Pelachaud, "From Non-verbal Signals Sequence Mining to Bayesian Networks for Interpersonal Attitudes Expression," in *Proceeding of the 14th International Conference on Intelligent Virtual Agents*, 2014, pp. 120–133.

[32] J. Broekens, C. Qu, and W. Brinkman, "Dynamic Facial Expression of Emotion Made Easy," 2012.

[33] C. Hu, M. A. Walker, M. Neff, and J. E. F. Tree, "storytelling agents with personality and adaptivity," in *Proceeding of the 15th International Conference on Intelligent Virtual Agents*, 2015, vol. 9238, pp. 181–193.

[34] R. Prada and A. Paiva, "Teaming up humans with autonomous synthetic characters," *Artificial Intelligence*, vol. 173, no. 1, pp. 80–103, Jan. 2009.

[35] N. Hussaini and R. Aylett, "Modelling Group Behaviour in Autonomous Synthetic Characters," in *15th International Conference on Intelligent virtual agents (IVA)*, 2015, vol. 9238, pp. 128–131.

[36] B. Ravenet, A. Cafaro, B. Biancardi, M. Ochs, C. Pelachaud, I. Mines-télécom, and T. Paristech, "Conversational Behavior Reflecting Interpersonal Attitudes in Small Group Interactions," in *15th International Conference on Intelligent virtual agents (IVA)*, 2015, pp. 1–14.

[37] S. Poeschl and N. Doering, "Designing virtual audiences for fear of public speaking training - An observation study on realistic nonverbal behavior," in *Annual Review of Cybertherapy and Telemedicine*, B. K. Wiederhold and G. Riva, Eds. Interactive Media Institute and IOS Press, 2012, pp. 218–222.

[38] A.-D. Tudor, S. Poeschl, and N. Doering, "What do audiencences do when they sit and listen?," in *Annual Review of Cybertherapy and Telemedicine*, 2013, pp. 120–124.

[39] N. Morina, W.-P. Brinkman, D. Hartanto, and P. M. G. Emmelkamp, "Sense of presence and anxiety during virtual social interactions between a human and virtual humans.," *PeerJ*, vol. 2, p. e337, Jan. 2014.

[40] C. Qu, Y. Ling, I. Heynderickx, and W.-P. Brinkman, "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*, vol. 10, no. 4, p. e0125279, Jan. 2015.

[41] M. Price, N. Mehta, E. B. Tone, and P. L. Anderson, "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*, vol. 25, no. 6, pp. 763–70, Aug. 2011.

[42] D. Hartanto, I. L. Kampmann, N. Morina, P. G. M. Emmelkamp, M. A. Neerincx, and W. P. Brinkman, "Controlling Social Stress in Virtual Reality Environments," *PLoS ONE*, vol. 9, no. 3, p. e92804, Mar. 2014.

[43] C. Qu, W. P. Brinkman, Y. Ling, P. Wiggers, and I. Heynderickx, "Conversations with a virtual human: Synthetic emotions and human responses," *Computers in Human Behavior*, vol. 34, no. May, pp. 58–68, May 2014.

[44] M. Slater, D.-P. Pertaub, C. Barker, and D. M. Clark, "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*, vol. 9, no. 5, pp. 627–33, Oct. 2006.

[45] C. Zanbaka, A. Ulinski, P. Goolkasian, and L. F. Hodges, "Social Responses to Virtual Humans: Implications for Future Interface Design," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2007, pp. 1561–1570.

[46] D. A. Clark and A. T. Beck, *Cognitive therapy of anxiety disorders: science and practice.* New York, New York, USA: The Guilford Press, 2010.

[47] J. Fox and J. N. Bailenson, "Virtual Self-Modeling: The Effects of Vicarious Reinforcement and Identification on Exercise Behaviors," *Media Psychology*, vol. 12, no. 1, pp. 1–25, Feb. 2009.

[48] D. H. Schunk and A. R. Hanson, "Peer models: Influence on children's self-efficacy and achievement.," *Journal of Educational Psychology*, vol. 77, no. 3, pp. 313–322, 1985.

[49] A. Bandura, *Self-efficacy: The exercise of control.* New York: W .H. Freeman and Company, 1997.

2

An Expressive Virtual Audience with Flexible Behavioral Styles

Currently, expressive virtual humans are used in psychological research, training, and psychotherapy. However, the behavior of these virtual humans is usually scripted and therefore cannot be modified freely at run time. To address this, we created a virtual audience with parameterized behavioral styles. This paper presents a parameterized audience model based on probabilistic models abstracted from the observation of real human audiences (n = 16). The audience's behavioral style is controlled by model parameters that define virtual humans' moods, attitudes, and personalities. Employing these parameters as predictors, the audience model significantly predicts audience behavior. To investigate if people can recognize the designed behavioral styles generated by this model, 12 audience styles were evaluated by two groups of participants. One group (n = 22) was asked to describe the virtual audience freely, and the other group (n = 22) was asked to rate the audiences on eight dimensions. The results indicated that people could recognize different audience attitudes and even perceive the different degrees of certain audience attitudes. In conclusion, the audience model can generate expressive behavior to show different attitudes by modulating model parameters.

This chapter has been published in IEEE Transactions on Affective Computing **4**, 326 (2013).

2

2.1. Introduction

Like a human audience, an audience of virtual humans has the ability to elicit responses in humans, e.g., [1], [2]. This ability makes a virtual audience beneficial when it comes to training, psychotherapy, or psychological stress testing. For example, it can help musicians to practice performing in front of an audience [3]. Virtual audiences are also being used as part of exposure therapy for individuals with social anxiety disorder [4] by exposing them to situations they fear. Instead of learning to cope with anxiety, some studies (e.g.,[5]) suggest that virtual audiences may also be used in the Trier Social Stress Test (TSST) [6] to induce stress in an individual with the aim of studying the effect of stress.

Besides the logistic advantage of not needing to arrange audience members and a suitable location, a virtual audience also offers the ability of control over the audience. For example, although the procedure for the standard TSST aims for a neutral audience, some have also explored variations with supportive or non-supportive audiences [7]. For exposure therapy, control of the fear stimuli is also desirable, as therapists aim to gradually expose patients to more fear-eliciting situations. Besides switching between different situations, e.g., an audience of fewer or more people [8], Emmelkamp [9] also suggests that treatment of social anxiety can also benefit from control over the fear stimuli within a virtual reality session, e.g., the behavior of the audience, as patients need to experience a certain amount of anxiety. Some treatment manuals [8] even give specific instructions on the desired anxiety level. Therefore, these manuals [8], [10] suggest using the attitude of an audience (e.g., negative or positive audience) as an effective means of controlling anxiety in a public speaking scenario. Currently virtual audiences are often represented by 3D models animated by a predefined script, e.g., [2], or videos of actual people embedded in Virtual Environments (VE)s, e.g., [11]. To control the audience's behavior, different animations or videos should be prepared so that operators can switch between these clips. However, the preparations may require considerable effort and thus are usually made in advance because explicit behaviors need to be scripted along the timeline for each audience member. Due to the effort involved, these prescripted animations and videos are often relatively short, causing the virtual audience to behave in repeating loops. This repetition may reduce the behavioral realism, thereby lowering the desired effect, e.g., lowering treatment efficacy [12]. From an engineering perspective, a more flexible and efficient system can be developed by applying software agents for the virtual audience to generate expressive behaviors automatically. Instead of specifying individual audience behavior, operators can adjust the agent parameters, e.g., attitude or personality, at run time to change the audience behavioral style. Controlling the audience on this higher level of abstraction reduces workload, as low-level audience behaviors no longer need to be controlled manually.

We therefore propose to use a statistical model, i.e., a model based on a corpus of audience behavior instead of theories of audience behaviors, to generate expressive behavior of virtual audience members. This method allows a human operator (e.g., researcher, therapist, or trainer) to control the virtual audience's behavioral styles by setting the agents' attributes (e.g., attitude and mood) and environment settings (e.g., interrupting events). This paper describes the creation of such a virtual audience, set within a public speaking scenario, as this is a commonly used laboratory procedure to elicit stress, e.g., as part of TSST, and as this is also one of the most common social situations that people with social anxiety fear [13]. Since the audience in public speaking situations usually shows their attitudes through body expressions, the design focuses on the generation of bodily responses of the virtual audience. To create such an audience, the main contributions of this study are: (1) a parameterized audience model which generates expressive behaviors based on statistical models, and (2) a corpus of audience behavior in public speaking situations.

2.2. DESIGN OF THE VIRTUAL AUDIENCE

As already mentioned, the behavior of the virtual audience should be realistic, flexible and expressive to display different attitudes. Thus, this paper proposes a parameterized agent model for the audience to generate expressive listening behavior. The behavioral styles can be modulated by adjusting agent attributes such as mood, attitude, personality, and energy level. The models for behavior generation are realized through a statistical approach.

2.2.1. REALISTIC AND FLEXIBLE EXPRESSIVE BEHAVIOR

Behavioral realism can be achieved by using autonomous agents. Although few studies have reported on the behavior of an autonomous audience, the potential for natural behavior has already been shown in recent studies of Embodied Conversational Agents (ECA). For example, a speaking agent can generate natural head movements [14], and a listening agent with simulated backchannel (head nod and smile) can improve the rapport in the human-agent interaction [15].

Adjustable expressive behavior can be implemented by a parameterized agent model. The parameters should affect the virtual humans' behavior so that they can behave expressively. For example, a model of listeners' feedback behavior in a multiparty conversation [16] was able to take into account several factors which affect agents' behavior, such as their conversational roles and goals, understanding, and attitudes. Furthermore, Busso et al. [17] shows the possibility of computational models to predict a speaker's head motion for different emotions, and their evaluations suggested that these models successfully emphasized the emotional content and improved the virtual speaker's behavioral naturalness.

Using such a parameterized model, operators can adjust the virtual audience's behavioral style by modifying its parameters. To convey affective connotations via body language, the parameters were selected from attributes that can affect and can be expressed in a person's nonverbal behavior. These attributes include moods, attitudes, personality, and physiological states [18].

2.2.2. BEHAVIOR GENERATION

Behavior generation of autonomous agents is often implemented by two main approaches: the theoretical and the statistical approach. The theoretical approach is to craft the rules that specify which behavior should be generated in a certain context based on psychological knowledge and literature. Examples using this approach include the listener model by Bevacqua et al. [19]. The statistical approach has also been widely used. It

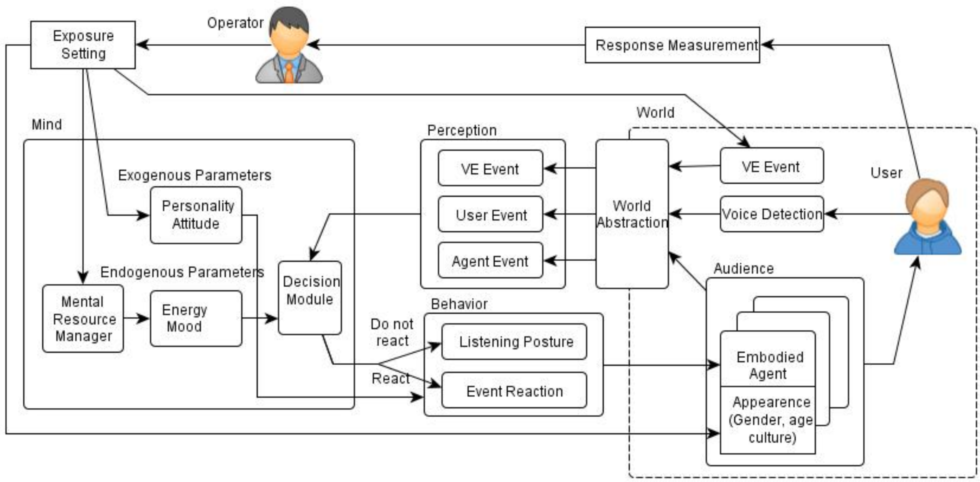


Figure 2.1: The framework of the virtual audience simulator. The arrows (→) in the diagram illustrate the direction of information flow.

uses statistical models taken from observations or corpora of human behavior to predict virtual agents' behavior. For example, a speaking agent [14] was developed using a machine learning approach, and listener's backchannel behavior (head nod and smile) [15] was generated by a probabilistic prediction model. Whereas the statistical approach needs real-life observations of a certain phe-nomenon to build a model, the theoretical approach requires more broad and general knowledge of the phenomena. At this moment, complete, coherent, and formal specifications of audience behavior cannot be derived from current theories; hence the statistical approach was applied in this study to generate the virtual audience's behavior.

2.3. SYSTEM OVERVIEW

THIS section describes the high-level design of the audience model based on the implementation methods discussed in Section 2.1. Figure 2.1 illustrates the framework of the integrated system and the architecture of the members of the autonomous audience. The overall structure of the agent architecture is based on common components of autonomous agents that should be able to perceive and act in the environment in which they are operating (see [20]). That is, the agent model includes a mind module for making decisions, a behavior module for translating the input from the mind into actions in the VE, and a perception module for perceiving the world (consisting of the VE and the user). Percepts work as input for the agent's decision making. In this way, the architecture implements a sense-reason-act cycle. This structure has also been widely used in ECAs, e.g., [15], [19]. Each module of this system is described in more detail later on, illustrating how they are composed to target the application of generating virtual audience behavior.

The *mind module* stores the values of the agent attributes. These attributes affect agent behavior and can be set by the operator. The agent attributes comprising personality, attitude (i.e., whether the agent is interested and positive towards the speech), mood, and energy level are assigned to two categories: the exogenous parameters and the endogenous parameters. Attributes regarded as static factors in a scenario such as personality and attitude belong to the exogenous parameters. These parameters can be set directly by the operator and remain constant unless they are modified by the operator. Dynamic attributes such as energy and mood belong to the endogenous parameters. As some ECA studies show, humans can perceive [21] and be affected by [22] virtual humans' affective states. Manipulation of the dynamic factors over time may be needed to regulate the user's state. Therefore, the endogenous parameters not only need initial values but are also influenced by the agent's *mental resource manager*. The mental resource manager stores emotional and physiological models, which can be defined by the operator and specify how the parameters change over time. According to the setting, the parameters will change automatically during the audience simulation. The parameters then feed into the *decision module*. Together with the perceived events such as a phone ringing or a fellow virtual human talking, the decision module will decide whether or not to react to these events. When the decision is made, the decision module will pass on the parameters to the *behavior module* to generate behavior.

The behavior module has two sub-modules: a *listening posture module* and an *event reaction module*. If the agent decides to react to a perceived event, it will output this decision to the event-reaction sub-module of the behavior module. The event-reaction sub-module will then generate an event response. If it decides not to respond or no event is perceived, it will directly pass the parameter values to the listening posture sub-module of the behavior module to generate a listening posture. The behavior module will generate a posture or movement every two seconds using one of its sub-modules (i.e., event reaction or listening posture). This posture or movement will then be used to animate the embodied agent in the VE.

While the user is giving a talk and the virtual agents are being animated, the agent perceives the *world* through the *perception module*. The world consists of the VE and the user, who is immersed in the VE. The perception module acquires the information from the *world abstraction module*, which works as an interface between the world and the agent. It provides abstracted information about the world such as a door slam (*VE event*), agent-agent interactions (*agent event*), and *user events*, e.g., a user's performance obtained by evaluating the user's speech using voice detection technology [23]. These events are passed on as percepts to the mind module, which can use the event information in its decision making. For example, when an agent perceives a door slam, it can decide to turn around and look at the door.

Besides the perception-mind-behavior model in this system, the operator has direct control over certain aspects of the VE, in particular, the appearances of the virtual agents and the occurrence of VE events to meet the audience simulation requirements.

2.4. Data Collection for the Agent Model

As the autonomous audience's behavior was generated based on statistical models abstracted from real-life observations, we observed and analyzed real audiences' behavior in different conditions. Pertaub et al. [24] found that the speakers' anxiety levels differ when they faced respectively a neutral static audience, a positive audience, and a negative audience (which exhibited bored and hostile expressions). To create an audience with more flexibility, besides the positive and neutral audiences, our observation data included two additional negative types: a critical audience and a bored audience. The critical audience was concerned about the speech topic but also critical of the talk. The bored audience was impatient and tired due to the boring speech. In summary, the virtual audience was designed to show at least four attitudes: positive, neutral, bored, and critical. Additionally, audiences' personality and mood data were included in the model to add realism and variety. Personality affects listening behavior [25] and can be perceived from virtual human's behavior (e.g., [19]). Studies (e.g., [26]) have also shown that mood can be expressed and perceived in several ways, e.g., postures and facial expressions. To achieve such a design, real audience's behavior was observed, and data of their personality, mood, attitude, and energy level was collected in four conditions: positive presentation, neutral presentation, boring presentation, and critical presentation. The audience behavior was then coded for the preparation of modeling the behavioral patterns.

2.4.1. Observation

Measures

The following measures were used to assess the personality, mood, attitudes, and energy levels of each audience member.

International Personality Item Pool (IPIP-NEO [27]). The IPIP-NEO is a public domain collection of items for per-sonality tests of adults. This study used a short inventory containing 120 items measuring the Big Five personality traits: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience. Each trait is scored on a scale of 0 to 99.

Self-Assessment Manikin (SAM [28]). SAM is a nonverbal pictorial assessment technique that directly measures three dimensions of mood: valence, arousal, and dominance. Each dimension has a 5-point rating scale.

Measure of Energy Levels (ME) and Measure of Attitudes (MA). Self-designed questionnaires were used to assess the audience members' energy levels and attitudes towards the presentations. The items of ME and MA were all rated on a scale of seven points, specified in Table 2.1.

Procedure

16 participants (seven females, nine males) were recruited from fellow PhD students studying computer science. The participants' ages ranged from 24 to 33 years ($M = 27.6$, $SD = 2.9$). All the participants signed the informed consent form. None of them knew the speaker beforehand. The participants were split into two eight-person groups. The behavior of participants in one group was video-recorded while they acted as an audience listening to four different presentations. Right before the first presentation, the audience was asked to complete the SAM questionnaire to assess their emotional states and the ME for energy levels. After each presentation, lasting around seven minutes, they were asked again to fill in the SAM and ME to track their emotional and energy states, and MA to acquire audience's attitudes towards the presentation. In each presentation, an interrupting event (i.e., a door slam or a telephone ring) was arranged randomly.

Table 2.1: Questionnaire items for measuring energy level and audience attitudes

Questionnaire		Item	Label 0	Label 6
ME	Energy	How is your current physical state?	Tired	Energetic
MA	Interest	What do you think of the topic?	Boring	Interesting
	Approval	How is your attitude towards the content?	Negative	Positive
	Eagerness for information	I was eager to get information and remember facts from the speech.	Extremely disagree	Extremely agree
	Criticism	I was critical to the speech and wanted to find flaws.	Extremely disagree	Extremely agree
	Impatience	I was impatient and hoped to finish as soon as possible during the speech.	Extremely disagree	Extremely agree

The whole process was repeated with the other group of participants, but this time the presentations were given in a reverted order to avoid the potential order effects. Thus, a set of videos consisting of four conditions for 16 participants was obtained. All 16 participants completed the IPIP-NEO personality inventory afterwards. Ethical approval for this study was obtained from the university ethics committee.

MATERIALS

The settings of the four presentations designed to evoke the four attitudes were as follows.

Positive presentation. To obtain the audience's interest, the audience was told at the beginning that they would win a small prize if they listened carefully and got a high score in the quiz afterwards. The topic was a novel invention of a robot gripper which was much more advantageous than traditional ones to evoke a positive attitude.

Neutral presentation. The topic was a software design method and there were no additional instructions for the audience.

Boring presentation. The speaker read aloud some text from the book of Nicomachean Ethics by Aristotle. However, the order of paragraphs had been rearranged so that the talk no longer contained a clear story line and therefore was no longer understandable for the listeners.

Critical presentation. The presentation criticized all the PhD students in the audience, saying that they did worst comparing with PhD students in other departments and in computer science departments at other universities. To address this, statistics were shown that only four out of 108 PhDs graduated on time over the last eight years and the average time needed for a PhD in the department to graduate was 5.5 years, which was a half year more than the average time needed for a PhD in computer science to graduate in the Netherlands. Additionally, a number of provocative policies were argued for, e.g., working hours from 9:00am to 6:00pm with only a half hour for lunch, and salary reduc-

Table 2.2: MA questionnaire results, Mean±SD

Questionnaire item	Presentation condition			
	Positive	Critical	Boring	Neutral
Interest	$5.06{\pm}1.18^{H}$	$4.19{\pm}1.68^{H}$	$0.69{\pm}1.08^{L}$	3.94±1.34
Approval	$4.94{\pm}0.77^{H}$	$2.44{\pm}1.86^{L}$	$1.44{\pm}1.75^{L}$	4.50±1.03
Eagerness for information	$4.75{\pm}1.61^{H}$	$4.50{\pm}1.37^{H}$	$1.88{\pm}1.71^{L}$	3.00±1.63
Criticism	$0.75{\pm}1.00^{L}$	$3.50{\pm}2.37^{H}$	$0.88{\pm}1.93^{L}$	1.63± 1.67
Impatience	$1.00{\pm}1.03^{L}$	$0.75{\pm}1.24^{L}$	$4.12{\pm}2.13^{H}$	1.25±1.24

Note: a mean with H indication is significantly (p < 0.01) higher than a mean with L indication within one questionnaire item.

tion if the research progress was slow. The speaker said that this presentation would also be given to the head of the department.

CONDITION VERIFICATION

To confirm that the audience attitudes were respectively positive, neutral, bored, or critical towards the presentations, the MA questionnaire results reflecting their attitudes were analyzed. The results (Table 2.2) show that the positive audience was significantly more interested ($t(15) = 10.02$, $p < 0.001$), more positive (i.e. high ratings in Approval, $t(15) = 7.12$, $p < 0.001$), more eager to get information ($t(15) = 4.37$, $p = 0.001$), and less impatient ($t(15) = -7.01$, $p < 0.001$) than the bored audience. The critical condition was similar to the positive condition but significantly less positive ($t(15) = -5.00$, $p < 0.001$) and more critical ($t(15) = 4.79$, $p < 0.001$) than the positive condition. The questionnaire results of the neutral condition were always between the high-level and low-level results. Therefore, the audience was respectively positive, neutral, bored, and critical in the corresponding conditions.

The questionnaire results were used as agent attributes in the agent modeling. As the attributes were rated on different scales (Table 2.3), all the attribute data was normalized by subtracting the minimum of each attribute from the raw value and then dividing the difference by the difference between the maximum and the minimum of the raw value. Hence the data for each attribute covered the whole scale. This result was then multiplied by 10 so that these attributes were rated on a common scale ranging from 0 to 10.

2.4.2. CODING POSTURES

To annotate the recorded video and characterize the audience's behavioral patterns, a posture-coding scheme was developed. The coding scheme describes how a certain part of the body moves in three-dimensional space using both anatomical and external reference frames [29]. A posture code consists of five sub-codes to convey position or movement information of the following parts: head, gaze, arms and hands, torso, and legs. Table 2.4 shows the posture-coding scheme for each part. That is, a sitting posture can be noted as a combination of the five sub-codes, i.e., HxGxAxTxLx. For example, if a person turns the head to the right (H2) and looks at the right side (G4), sitting up

Table 2.3: Statistics of audience data used as agent attributes

Dimension/ measures	Parameters	Measuring scale	Raw data			
			Min	Max	Mean	SD
Mood/ SAM	Valence	1 - 5	1	5	2.46	0.93
	Arousal	1 - 5	1	5	3.55	1.01
	Dominance	1 - 5	1	5	2.86	0.92
Energy/ME	Energy	0 - 6	0	6	3.61	1.38
Attitude/ MA	Interest	0 - 6	0	6	3.47	2.12
	Approval	0 - 6	0	6	3.58	1.99
	Eagerness for information	0 - 6	0	6	3.53	1.94
	Criticism	0 - 6	0	6	1.69	2.09
	Impatience	0 - 6	0	6	1.78	1.99
Personality/ IPIP-NEO	Openness to Experience	0 - 99	0	77	36.81	24.02
	Conscientiousness	0 - 99	1	99	60.56	29.78
	Extraversion	0 - 99	0	96	46.00	28.08
	Agreeableness	0 - 99	7	88	58.13	25.41
	Neuroticism	0 - 99	1	64	28.50	20.37

Figure 2.2: A video screenshot of an observed audience.

straight (T1) with a hand tapping on the desk (A17) and twisted ankles (L2), the posture will be annotated as H2G4A17T1L2. Although the coding scheme includes the audience's gaze information, the gaze has not been implemented in the behavioral model currently. Hence the postures that are mentioned below are combinations of positions or movements of the four parts: head, arms and hands, torso, and legs.

Figure 2.2 shows a video screenshot of an observed audience. To determine the sampling interval [30], the codable behaviors in the videos were analyzed. The shortest duration of a codable state was 2 seconds. Thus, with an interval of two seconds, a coder coded each audience member's postures with the posture-coding scheme. The postures were coded by recording four position variables: head, arms and hands, torso, and legs. Taking the coding of leg positions for example, the coder had a choice out of three position codes (i.e., L1, L2, and L3) for each coding unit. Moreover, when an interrupting event occurred, additional information was recorded, specifically the reaction of each participant (i.e., turning angle of the head and reaction duration) and the event infor-

Table 2.4: The posture coding scheme

Part of body	Behavior category	Short description
Head	H1: Head up	The head keeps the neutral position.
	H2: Head turn	The head turns right or left.
	H3: Head down	The head is lowered.
	H4: Head tilt	The head tilts right or left.
	H5: Head nod	Nod the head: the head moves down and then up again quickly.
	H6: Head shake	Shake the head: the head turns from side to side.
Gaze	G1: Towards the speaker	The gaze is directed towards the speaker.
	G2: Upwards	The gaze is directed above the speaker position.
	G3: Downwards	The gaze is directed below the speaker position.
	G4: Averted sideways	The gaze is directed away from the speaker
Arms and hands	A01: Hands on legs	Both hands are on the legs.
	A02: Open arms on desk	Both arms rest on the desk without touching each other.
	A03: Arms crossed	The arms are crossed in front of the body.
	A04: Catapult	The hands are holding behind the head like a catapult.
	A05: Hands steeple	The fingers of one hand lightly press against those of the other hand to form a church steeple.
	A06: Hands clenched	The hands are clenched, and the elbows rest on the desk.
	A07: Chin/ cheek touch	One hand touches the chin or cheek, and the other arm rests in front of the body.
	A08: Supporting head	One or two arms support the head and the elbows rest on the desk.
	A09: Desk and chair back	One arm rests on the desk and the other rests on the chair back.
	A10: Self-hold	One arm swings across the body to hold or touch the other arm.
	A11: Desk and leg	One arm rests on the desk and the other rests on one leg.
	A12: Nose touch	One hand touches the nose, and the other arm rests on the front torso.
	A13: Eye rub	One hand rubs the eyes, and the other arm rests on the front torso.
	A14: Ear touch	One hand touches the ear, and the other arm rests on the front torso.
	A15: Neck touch	One hand touches the neck, and the other arm rests on the front torso.
	A16: Mouth touch	One hand touches the mouth, and the other arm rests on the front torso.
	A17: Hand tap	One or two hands tap the desk continuously.
Torso	T1: Torso upright	The torso keeps upright.
	T2: Torso forward	The torso leans forward and the spine keeps straight.
	T3: Torso backward	The torso leans backward and the spine keeps straight.
	T4: Torso back in the chair	The torso leans back in the chair and the spine is relaxed.
	T5: Torso bent forward	The torso leans forward, and the spine is bent forward.
Legs	L1: Standard position	The knees are bent at a right angle with both feet flat on the ground, and the legs are not crossed.
	L2: Legs crossed/ twisted	The legs are crossed or the ankles are twisted.
	L3: Leg joggle/ tap	The upper leg joggles or the lower leg taps the floor when the legs are crossed, or one or two feet tap the floor when both feet rest flat on the floor.

mation (i.e., the direction and the duration of the event). To assess the coding reliability, an additional coder was trained and independently coded an eight-minute video sample of one audience member according to the coding scheme. The sample consisting of 240 units was coded out of a total length of 320 minutes (i.e., 9600 units). To avoid a biased sample, a representative sample was selected with similar frequencies of posture shifts as observed on average in the whole corpus, i.e., 1.25, 1.50, 0.63, and 0.38 behavior shifts per minute for head, arms and hands, torso, and legs respectively for the sample, and 1.26 ($SD = 1.66$), 1.19 ($SD = 1.27$), 0.40 ($SD = 0.76$), and 0.34 ($SD = 0.88$) for the whole corpus. Like the first coder, the second coder coded the four position variables at two-second intervals from the same starting point. The coding agreement between the two coders was assessed by computing Cohen's kappa [31] for each variable. The agreement coefficients for the four position variables were respectively 0.85, 0.85, 0.94, and 0.93, which shows an acceptable level of agreement [32]. Note that the combination of relatively few behavioral shifts and the relatively short two-second sampling intervals created large sequences without variations, which might cause relatively high agreement level.

From the recorded videos lasting around 320 minutes, over 300 unique postures (presented by unique combinations of the four sub-codes, HxAxTxLx) were observed. To simplify the analysis and system implementation, the postures which occurred less than 6 times (an occurrence of a posture was counted only when the posture changed to another) in the whole observation were removed, resulting in 59 postures. The remaining coding still accounted for 80% of the 9600 coding units.

The collected data was used to build statistical models of audience behavior. The next two sections explain the statistical models and how the *mind module* and the *behavior module* use these models to generate the virtual agent's behavior.

2.5. THE MIND MODULE

THE *mind module* stores the agent attributes that affect the virtual agent's behavioral style and includes a *decision module* for event response. The agent attributes, including personality, attitude, energy and mood, are presented by parameters listed in Table 2.3. Before passing the parameters to the *behavior module*, the *decision module* will first decide whether or not the agent should respond in case of interrupting events.

To mimic the probabilities with which real persons would respond to the events, the decision model was trained using the observed data. A supervised classification method, logistic regression, was applied. The agent parameters (Table 2.3) and event information (i.e., event duration and event location) were used as predictors. The training and test data used the normalized questionnaire results and the coding information of the observations. For the endogenous parameters, such as energy level and mood, data was collected before and after each presentation. To simplify the model, the values of these parameters were assumed to change linearly during the presentation. Hence the missing data during each talk for mood and energy level were linearly interpolated using the results of ME and SAM before and after each presentation.

A sample of 39 cases was drawn from the original data set with an almost equal number of cases where a person did or did not respond to an event. This avoided a biased function caused by an imbalanced data set where the sizes of classes are not similar. The logistic regression model can be expressed by the following formulae:

$$p = 1/1 + exp[-(b_0 + b_1x_1 + b_2x_2 + ... + b_kx_k)]$$
$$Y = \begin{cases} 1 & p \geq 0.5 \\ 0 & p < 0.5 \end{cases} \tag{2.1}$$

where b_0, b_1, b_2, ..., and b_k are regression coefficients for predictor variables, x_1, x_2, ..., and x_k, p presents the probability for the agent to respond, and Y is the prediction output. The model selected a cutoff point of 0.5, i.e., the prediction is to respond to events ($Y = 1$) when p is no less than 0.5, otherwise the prediction is not to respond ($Y = 0$). A test of the full model versus a model with intercept only was statistically significant, $\chi^2(3, N = 39) = 18.92$, $p < 0.001$, with an overall correct prediction of 81.2% (85.0% for non-response and 78.9% for response cases). This model was also tested on 11 cases (five response and six non-response cases) that had not been used for training. The overall correct prediction was 90.9% (i.e., five response and five non-response cases were correctly classified). This result was significantly (binomial test, $p = 0.01$) above a case allocation of 54.54% (i.e., 6 out of 11 cases).

2.6. The Behavior Module

The *behavior module* generates listening postures using the parameters from the *mind module* and generates head turns as an event reaction if the mind has decided to react.

2.6.1. Generation of Listening Postures

To derive listening postures from agent parameters, a relationship between the parameters and listening behavior was established in the module. To do this, the 59 observed postures (see Section 2.4.2) were first categorized so that the agent attributes could be used to predict a category, from which a posture would be selected afterwards.

The 59 postures were clustered according to their transition probabilities to other postures. The posture sequences were transformed to a 59×59 transition matrix, i.e.,

$$P = \begin{bmatrix} p_{1,1} & \cdots & p_{1,j} & \cdots & p_{1,59} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{i,1} & \cdots & p_{i,j} & \cdots & p_{i,59} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{59,1} & \cdots & p_{59,j} & \cdots & p_{59,59} \end{bmatrix} \tag{2.2}$$

where $p_{i,j}$ is the probability of transitioning from posture i to posture j within two successive observations, i.e., every two seconds. The postures with similar transition probabilities were clustered into one category. Take Table 2.5 for example, each row presents the probabilities for one posture to transition to Posture 1, 2, and 3 respectively. As the probabilities for Posture 1 and 2 are very similar, i.e., all around 0.71, 0.29, and 0.00, the two postures were clustered together. This also means that the postures in one category were always followed by postures from a certain posture set (here in this example the set consists of Posture 1 and 2). Therefore, each row of the transition matrix shown in (2),

Table 2.5: An example of the posture transition matrix with accumulative probability intervals

		Category 1		Category 2
		Posture 1	Posture 2	Posture 3
Category 1	Posture 1	0.73	0.27	0.00
		[0, 0.73)	[0.73, 1]	(1, 1]
	Posture 2	0.69	0.31	0.00
		[0, 0.69)	[0.69, 1]	(1, 1]
Category 2	Posture 3	0.02	0.03	0.95
		[0, 0.02)	[0.02, 0.05)	[0.05, 1]

presenting the probabilities of transition from one posture to others, was used as a clustering feature. An agglomerative hierarchical clustering method with Ward linkage was then employed to group the postures using a Euclidean distance measurement. A distance threshold of 20 on a scale from 0 to 25 was set to seek an optimum in the maximum number of categories and a maximum number of similar postures within each category. Thus, 15 categories were identified with each containing 3 to 7 similar postures, which often differ only in one of four coding parts, e.g., H1A08T5L3 and H4A08T5L3.

The second step was to create a logistic regression prediction model using agent parameters (i.e., mood, energy, attitude, and personality) as predictors. Since logistic regression predicts a dichotomous outcome, i.e., whether the parameter set belongs to a certain category or not, the models were trained separately to predict each category. Like the training data for event reaction, this training data was also randomly sampled so that the data for each category was distributed equally. 15 prediction functions were established using the following form:

$$p_i = 1/1 + exp[-(b_{0i} + b_{1i}x_1 + b_{2i}x_2 + ... + b_{ki}x_k)] \tag{2.3}$$

where p_i represents the probability of being category i, x_1, x_2, ..., and x_k are the predictor variables, and b_{0i}, b_{1i}, ..., and b_{ki} are the regression coefficients for category i. To predict the exact category using a set of agent attributes, x_1, x_2, ..., and x_k, one probability was calculated for each of the 15 categories by (3) respectively, and eventually based on this set of probabilities the predicted category n would be selected that satisfies

$$p_n = max\{p_1, p_2, \cdots, p_{15}\} \tag{2.4}$$

The overall correct prediction of the training set was 66.4%, ranging from 48.2% to 88.6% for individual categories.

As 20% of the balanced observed data had not been used for training, it was possible to conduct a holdout validation for the model. The test data included 50 cases of each category, i.e., 750 cases in total. The results of the test set showed an overall correct prediction of 64.4% for the 15 categories, which was significantly (binomial test, $p < 0.001$) above the random allocation threshold of 6.7% (i.e. 1 out of 15). For the 15 individual categories, correct prediction ranged from 48.0% to 94.0%, which were all significantly (binomial test, $p < 0.001$) above the random allocation threshold of 6.7%. Interestingly, we compared the effect of exogenous parameters (i.e., personality and attitude) and endogenous parameters (i.e., energy and mood) on the prediction of a category by analyzing the odds ratios of the regression coefficients, i.e., $\exp(b_{ki})$. The absolute value of

b_{ki} was examined so that odds ratios exp(b_{ki}) and exp(-b_{ki}) can reflect the same impact on the category selection. For each function, the sums of the absolute coefficients for exogenous and endogenous parameters were calculated, i.e., and where m and n are the numbers of exogenous and endogenous parameters in the ith function. By comparing the sums of all 15 functions, the effect of exogenous parameters ($M_{\sum|b(exo)|}$ = 19.22, $SD_{\sum|b(exo)|}$ = 25.32) was found to be greater than that of endogenous ones ($M_{\sum|b(end)|}$ = 1.98, $SD_{\sum|b(end)|}$ = 2.82), $t(14) = -2.87$, $p = 0.01$. Compared with the endogenous parameters, the exogenous had more effect on the category selection on average. Still all parameters contributed to the model significantly ($p < 0.05$) according to the Wald statistics.

After a behavior category is determined, a posture will be selected within this category. Since the *behavior module* updates the embodied agents' behavior every two seconds, the category is very likely to remain unchanged. When the category does not change, the posture will be selected according to the transition matrix to keep the sequential pattern of the behavior. For example, supposing that the current posture category is 1 and the current posture code is 1, the following posture should be selected within Category 1. According to the transition probabilities in Table 2.5, the next posture has a chance of 0.73 to be Posture 1 and 0.27 to be Posture 2. To select a posture, a random number between 0 and 1.00 is generated as the accumulative probability. If the number is 0.96, which is within the range of [0.73, 1], Posture 2 will be selected.

2.6.2. Generation of Event Responses

Since turning one's head was the only event response considered in the observation data, the module only determines how many degrees an agent turns its head (turning angle TA) and how many seconds the response takes (response duration DUR). In the audience observation, information of interrupting events and audience responses was recorded. This information as well as the audience attributes was used to predict the turning angle of a virtual human's head TA and its response duration DUR.

Linear regression results indicated that only the event direction, which is a degrees ($-180° < a \leq 180°$) relative to the front direction of the virtual agent, significantly predicted the agent's turning angle TA ($R^2 = 0.78$, $F(1, 18) = 63.03$, $\beta = 0.88$, $p < 0.01$) by the following model:

$$TA = 0.517a \tag{2.5}$$

The event direction also significantly predicted the agent's response duration DUR in seconds ($R^2 = 0.80$, $F(1, 18) = 73.65$, $\beta = 0.90$, $p < 0.01$), in the following form:

$$DUR = 0.013|a| \tag{2.6}$$

As indicated by the two models, when an event occurs right behind the virtual human (i.e., $a = 180°$), the virtual human would turn its head about 90 degrees, and the response duration lasts about 2.3 seconds at maximum. Contrarily, if an event is in front of the agent, assuming that $a < 20°$, the reaction duration would be very short according to the rules, i.e. $DUR < 0.26$s. In practice, this short head turn would not be acted out, which coincides with our observation: when an event occurs in front, the head turn was unnecessary because the event was still in the audience member's field of view.

2.7. The World and the Perception Module

As explained in Section 3, the agent perceives the world through the *perception module*. The information about what is happening in the world needs to be abstracted into events to be usable for the *decision module*. The interrupting events in the *world* include *VE events*, *agent events*, and *user events*. The *VE event* is evoked by the "physical" objects in the VE, such as door slam and telephone ring, and can be set in the therapy settings. The *agent event* refers to the agent's behavior which may evoke interaction with another agent. For example, a head turn is generated by the *behavior module* so that an agent look at another agent for a while, which may cause another agent to turn the head back. The *user event* relates to the user's performance which may evoke the agents to change their behavior, e.g., the user stops talking for a moment, which may result in the distraction of the agents.

The way in which information is abstracted depends on what is needed at the level of decision making. For example, concerning the user speaking, it may be enough to generate an event to indicate whether she is speaking or not. However, if this system enables speech interaction with the human speaker, e.g., asking questions about the talk, the *world abstraction module* needs to pass on more detailed information about it, such as the topic and key words. Thus, the *world abstraction module* also works as an information provider and makes the system easy to be extended.

While the agents perceive the world, the operator should be aware of the information from the world too. The user's response feedback may include information such as gaze direction [33] and anxiety or stress level by measuring anxiety such as subjective unit of discomfort and psychophysiological data [34]. Additionally, information on the interrupting events in the VE could be recorded with response measurement in a log file so that the user's response can be analyzed afterwards.

Moreover, the operator has direct control over certain aspects of the VE, e.g., the virtual human's appearances which can be determined by static appearance parameters like gender and age, and the occurrence of VE events defined by parameters such as event location and event duration. Other controllable appearance elements could also be added such as ethnicity and clothes to construct a more realistic environment [35].

2.8. Perception Evaluation of a Virtual Audience

This study proposes a framework for a public speaking simulation system in which an operator can control the behavioral styles of an autonomous virtual audience. Among the components of this system, this study mainly focused on the audience model and the creation of such an audience. Since the previous sections already show that the audience model fits well with the corpus data, a next step was to examine how people perceive the audience. To do this, an autonomous audience in a public speaking situation was created using this model so that individuals could evaluate the model by watching the audience's behavior.

2

2.8.1. Method

Hypotheses and Experiment Design

The hypotheses for this evaluation were that people could perceive the different audience's attitudes (H1), moods (H2), and personalities (H3) from the behavioural styles modulated by corresponding parameters.

To test H1, the evaluation mainly examined the perception of the four designed attitudes: critical, positive, neutral, and bored. To further investigate whether people can recognize the different degrees of a certain attitude, the positive and bored attitudes respectively included two conditions: an extremely positive condition and a positive condition, and, an extremely bored condition and a bored condition. Hence there were six attitude conditions: a critical attitude, an extremely positive attitude, a positive attitude, a neutral attitude, a bored attitude, and an extremely bored attitude.

Concerning H2 and H3, the study only explored some of the mood and personality dimensions, namely, valence, arousal, and extraversion. These dimensions were selected because these dimensions may be perceived more easily than others, e.g., neuroticism [36]. Thus, the study also includes six additional audience conditions labeled as follows: extrovert, introvert, high arousal, low arousal, positive valence, and negative valence.

To test these hypotheses, the evaluation was conducted in two ways with different participants. One group of participants was asked to describe the audience's state freely. This open-question avoided framing their observations or biasing the participants' response towards a specific factor. Therefore, the audience description should reflect their natural thoughts. The second group was asked to rate their observation with a questionnaire to obtain information on the factors to be examined.

Materials

An executable program was made to display the simulation of a public speaking situation in which a 12-person audience was seated in a classroom. The executable program generated the audience's behavior in real time so that participants watched different audience animations due to the random element in the simulation. The viewpoint was set from the perspective of an outsider in front of this audience slightly on the right (Figure 2.3). The speech the audience listened to was selected from the news report in Uygur language so that participants could not understand the speech and therefore would not be affected by the speech content.

12 audience conditions of one minute each were created to show different attitudes, moods, and personalities. These 12 conditions were created by setting each agent's attributes as Table 2.6. The attitude conditions were made by modulating the attitude parameters. Specifically, the four conditions, namely critical, extremely positive, neutral, and extremely bored, were created according to the results of attitude questionnaire (MA) obtained in the four observed situations (see Table 2.2). The positive and the bored conditions employed moderated settings of the extremely positive and the extremely bored conditions by adjusting one or two parameters from the extremities to a medium level. Figure 2.3 shows the snapshots of the audience in six attitude conditions.

Instead of using the observed audience conditions, the effects of the mood and personality parameters were explored by setting extremities in the examined dimensions,

(a) Neutral

(b) Critical

(c) Positive

(d) Extremely positive

(e) Bored

(f) Extremely bored

Figure 2.3: Snapshots of the autonomous audience in six attitude conditions.

e.g., the extrovert condition only set the parameter Extraversion as high.

Measures

A questionnaire about the virtual audience was designed to quantitatively measure the perceived audience's attributes, including attitude, mood, and personality. The attitude and mood questions were adapted from MA and SAM (section 2.4.1 Measures) to refer to the virtual audience's state. For example, the criticism item in the questionnaire became:

> *The audience was critical towards the speech and wanted to find flaws.*

Additionally, the questionnaire did not include the dimensions that were not evaluated in the 12 conditions, e.g., the dominance item in SAM. Thus, the questionnaire included five attitude items and two mood items. It also included a personality item, formulated as follows:

> *Most audience members scored high on extraversion*

Table 2.6: Parameter settings for the audience conditions

Condition	Parameter setting
Critical	Attitude parameter: Interest (H), Approval (L), Eagerness for information (H), Criticism (H), Impatience (L).*
Extremely Positive	Attitude parameter: Interest (H), Approval (H), Eagerness for information (H), Criticism (L), Impatience (L).
Positive	Attitude parameter: Interest (H), Approval (H), Criticism (L).
Neutral	*This condition was set as baseline: all parameters were set at the medium level.*
Bored	Attitude parameter: Interest (L), Eagerness for information (L), Criticism (L), Impatience (H).
Extremely Bored	Attitude parameter: Interest (L), Approval (L), Eagerness for information (L), Criticism (L), Impatience (H).
Extrovert	Personality parameter: Extraversion (H).
Introvert	Personality parameter: Extraversion (L).
Positive Valence	Mood parameter: Valence (H).
Negative Valence	Mood parameter: Valence (L).
High Arousal	Mood parameter: Arousal (H).
Low Arousal	Mood parameter: Arousal (L).

** Each condition was set up by attitude (MA), mood (SAM), and personality (IPIP-NEO) parameters. This table only specifies the parameters which were set as High (9) or Low (1). All the other parameters were set at the neutral level (5).*

This item was rated on a 7-point scale, from “extremely disagree” (i.e., 0) to “extremely agree” (i.e., 6).

Therefore, the hypotheses could be tested by comparing these measurements with the parameter settings of the evaluation conditions.

PROCEDURE

The evaluation included two parts: free description and factor rating. For the first part, 22 participants (10 females, 12 males) were recruited throughout the university campus to evaluate the virtual audience. Their ages ranged from 22 to 38 years with a mean of 28.1 ($SD = 3.2$) years. Each participant was asked to watch a 12-minute audience simulation using a Sony HMZ-T1 head-mounted display (HMD) with an orientation tracker to track the participant’s head orientation. The HMD displayed a virtual image comparable to viewing a 720-inch display at 20 meters and the visual field spanned 45 degrees diagonally. The resolution of the right and left display was 1280*720 (horizontal*vertical) pixels with a refresh rate of 60Hz.

While watching the simulation, the participant was asked to observe and describe orally the state of the audience. To avoid framing or biasing the participants’ description, no examples of audience description were given to the participants. Their description was audio recorded. Among the 22 participants, 16 participants were Chinese and reported in standard Chinese in the experiment, and six other participants were Dutch and Iraqi and they reported in English. The order of those 12 conditions was randomly given to each participant to avoid the potential order effects.

For the factor rating part of the evaluation, another 22 participants (13 females, 9 males) were recruited. Their ages ranged from 23 to 44 years with a mean of 28.5 (*SD*

Table 2.7: The coding scheme for the recorded description and reliability assessment

Description category	Short description and utterance examples	Cohen's kappa
Attentive	The audience pay attention to the speech and may be attracted: "paying attention", "attentive", "interested", "thoughtful about the speech content", "curious", "concentrate", "focus", "pleased with the speech", "positive attitude"	0.94
Neutral	The audience do not show any interest or negative attitude towards the speech: "neutral", "neutrality"	*
Distracted	The audience do not pay attention to the speech: "distracted", "looking away", "inattentive", "day-dreaming"	0.88
Bored	The audience are bored with the speech and impatient to wait to the end: "bored and tapping the desk/ floor", "frustrated about being there", "impatient", "lose interest", "unhappy to be there"	0.87
Critical	The audience show disapproval towards the speech: "disagree", "critical", "try to find flaws", " angry with what is talking about", " skeptical"	1.00
Active	The audience are active and aware of the surroundings: "active", "awake", "conscious", "alert"	*
Sleepy	The audience almost fall asleep or show a state of low energy level: "sleepy", "slouching", "tired"	1.00
Relaxed	The audience are physically relaxed, instead of sitting straight up: "relaxed"	0.87

** No statistics were computed because no rater had coded the sample recordings using this category, resulting in two constant coding sequences.*

= 5.5) years. The participants included four therapists and four psychology master students who all had experience in using virtual reality exposure system [37] to treat patients with social anxiety disorder. The other 14 participants, with no such experience, were recruited throughout the university campus. Like the first part, each participant was asked to watch the 12 audience conditions using the HMD in a random order. However this time, the participants were asked to rate the factors with the questionnaire after watching each condition.

2.8.2. ANALYSIS AND RESULTS

FREE DESCRIPTION

To statistically investigate whether the participants could recognize the different conditions, a coding scheme for their description was developed, shown in Table 2.7. Each participant's comments for each minute were analyzed afterwards by a coder. The coder recorded whether or not the comments in a condition included terms that would fall into one or more of the eight description categories. In this way, a set of eight binary digits was obtained per participant per condition.

To assess the reliability of the coding, another coder was trained to code the audio recordings according to the coding definitions. The additional coder coded independently a sample of 36 minutes out of a total length of 264 minutes. The coding agreement between the two coders was assessed by computing the Cohen's kappa [31]. Table

Table 2.8: Number of participants who used certain description categories and results of Wilcoxon signed rank tests ($n = 22$)

Condition		Description category: Attentive	Neutral	Distracted	Bored	Critical	Active	Sleepy	Relaxed
	Neutral	12	1	16	8	2	1	5	2
Comparison with Neutral condition	Critical	11	0	7*	4	2	0	4	1
	Extremely Positive	18*	0	7*	4	0	2	5	4
	Positive	15	0	12	4	1	2	5	2
	Bored	8	0	9	12	1	1	3	2
	Extremely Bored	4*	0	16	15*	0	0	6	0
Comparison with its opposite condition	Extrovert	13	2	6	6	4	0	3	3
	vs Introvert	16	2	6	8	3	2	0	2
	Positive	14	2	3	10	2	3	3	1
	vs Negative Valence	13	0	9	7	4	1	4	3
	High	13	3	9	6	1	0	7	2
	vs Low Arousal	15	0	9	6	0	2	4	1

*$p < 0.05$

2.7 also presents the agreement coefficients varying from 0.87 to 1.00, showing an acceptable agreement level [32].

After coding all the recordings, 22 sets of binary data were obtained from all the participants for each condition. The data sets were then added up to count how many participants have mentioned a certain category in one condition (Table 2.8) to establish an overview of the differences across those conditions.

To examine if participants' utterance responses differed among the conditions, Wilcoxon signed rank tests for two related samples were conducted on each description category. The attitude conditions were respectively compared with the neutral condition, which was regarded as the baseline condition. Other conditions were respectively compared with their opposite conditions, e.g., Positive Valence versus Negative Valence. The results are also presented in Table 2.8.

The participants described the audience as distracted significantly (z = -2.32, p = 0.02) more often in the Neutral condition than the Critical condition. Furthermore, they described the audience as attentive more often (z = -2.12, p = 0.03) and as distracted less often (z = -2.32, p = 0.02) in the Extremely Positive condition than in the Neutral condition. Finally, compared with the Neutral condition, the audience in the Extremely Bored condition was described as less attentive (z = -2.14, p = 0.03) and more bored (z = -2.65, p = 0.008). This suggested that the participants could significantly differentiate an extremely positive or extremely bored audience from a neutral audience.

No significant difference was found in comparisons between Positive and Neutral

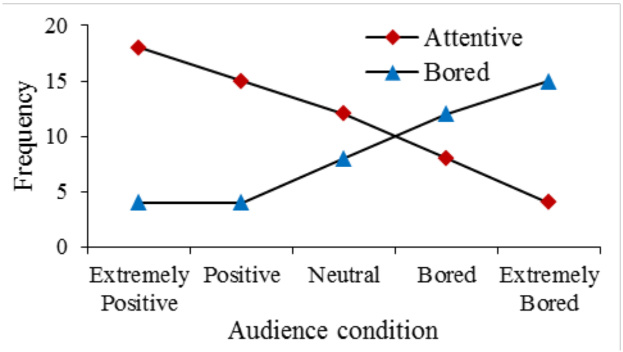


Figure 2.4: The frequency of Attentive and Bored in audience description against audience attitude conditions.

and between Bored and Neutral conditions. However, when positioning the description results in the order of Extremely Positive, Positive, Neutral, Bored, and Extremely Bored condition, a trend seems to appear for the Attentive and Bored categories, as shown in Figure 2.4. This suggested that the participants may even perceive the different degrees of a certain attitude, e.g., differentiating extremely positive attitude from positive attitude.

For the exploratory conditions, it seems that the participants did not make any reference with regard to the moods and personalities.

FACTOR RATING

Nonparametric tests were also conducted on rated items for the audience conditions because the ratings were not normally distributed. The analysis results of ratings are shown in Table 2.9. To investigate whether the perceptions of therapists and non-therapists were consistent with each other, Spearman's correlations were calculated between the medians of eight (students-)therapists and 14 non-therapists for the different items. The correlations ranged from weak positive (0.23) to very strong positive relationships (0.77) with an average strong positive relationship (0.52), which suggests a reasonable level of agreement between the two groups across the items. Therefore, the analysis was conducted on the data from all participants.

Table 2.9: Median and test results of factor ratings for therapists (T, n =8), non-therapists (NT, n =14), and all participants (A, n =22)

Condition	Questionnaire item: Interest			Criticism			Approval			Eagerness for information			Impatience			Valence			Arousal			Extraversion		
	NT	T	A	NT	T	A	NT	T	A	NT	T	A	NT	T	A	NT	T	A	NT	T	A	NT	T	A
Neutral	3.0	3.5	3.0	3.0	3.0	3.0	3.0	3.5	3.0	3.0	2.5	3.0	3.0	2.5	3.0	4.0	4.0	4.0	4.0	2.0	4.0	3.0	3.0	3.0
Critical	4.0	4.0	4.0^{H}	3.5	3.0	3.0^{H1}_{H2}	3.5	4.0	4.0^{H}	4.0	4.0	4.0^{H1}	3.0	2.5	3.0^{L}	4.0	4.0	4.0^{H}	3.5	2.5	3.0	3.0	3.5	3.0^{H}
Extremely Positive	4.5	4.0	4.0^{H}	3.0	1.0	1.5^{L2}	3.5	4.0	4.0^{H}	5.0	4.0	4.0^{H1}_{H2}	1.0	2.0	1.0^{L}	4.0	4.0	4.0^{H}	4.0	3.0	3.0	3.0	3.0	3.0^{H}
Positive	4.0	2.5	4.0^{H}	2.0	1.0	2.0	3.0	2.5	3.0	3.0	2.5	3.0	2.0	2.5	2.0^{L}	4.0	4.5	4.0^{H}	3.0	3.0	3.0	3.0	3.5	3.0^{H}
Bored	4.5	3.5	4.0	3.0	2.0	3.0	3.0	3.5	3.0	4.0	2.5	3.0^{L2}	2.5	2.5	2.5	3.5	4.0	4.0	3.5	2.0	2.0	3.0	3.0	3.0
Extremely Bored	1.0	0.5	1.0^{L}	1.5	1.0	1.0^{L1}	1.5	1.0	1.0^{L}	1.0	1.0	1.0^{L1}	4.5	5.0	5.0^{H}	2.0	2.5	2.0^{L}	2.0	1.5	2.0	2.0	1.5	2.0^{L}
Extrovert	4.0	3.0	4.0	2.5	2.0	2.0	3.5	3.0	3.0	4.0	3.0	4.0	2.0	2.5	2.0	4.0	4.0	4.0	4.0	2.0	2.5	4.0	3.5	4.0
Introvert	2.5	2.5	2.5	3.0	1.5	3.0	2.5	2.5	2.5	3.0	1.5	2.0	3.0	3.5	3.5	2.5	4.0	3.0	3.5	2.0	2.0	3.0	2.5	3.0
Positive Valence	5.0	4.5	5.0	3.5	1.0	3.0	4.0	4.0	4.0	5.0	3.5	4.5	1.5	2.5	2.0	4.0	4.0	4.0	4.0	3.0	4.0^{H}	3.0	3.5	3.0
Negative Valence	4.5	4.0	4.0	3.0	1.5	2.5	3.5	3.5	3.5	4.0	3.0	4.0	2.0	2.0	2.0	4.0	4.0	4.0	2.5	2.0	2.0^{L}	3.5	1.5	3.0
High Arousal	3.0	3.0	3.0	3.0	1.5	2.5	3.0	3.0	3.0	3.0	2.0	3.0	3.0	4.0	3.5	4.0	4.0	4.0	3.0	2.0	2.5	3.0	2.5	3.0
Low Arousal	4.0	2.5	4.0	4.0	2.0	3.0	3.5	2.0	3.0	4.0	1.5	4.0	1.5	3.5	2.0	4.0	3.0	4.0	4.0	2.0	2.0	3.0	3.0	3.0
Corrrelation between NT and T	0.76**			0.38			0.63*			0.77**			0.59*			0.40			0.36			0.23		

**p<0.05; **p<0.01.*

Note: a median with H or Hx indication is significantly ($p < .05$) higher than a median with L or Lx indication in the same column whereby indices refers to the specific pair which was compared. For example, the median of Eagerness for Information in the Critical condition was 4.0^{H1} which was significantly higher than 1.0^{L1}, the median in the Extremely Bored condition.

To verify whether the differences across the conditions correspond to the condition settings as hypothesized, all the conditions were compared with baseline conditions. The Extremely Bored condition was selected as the baseline for attitude conditions, hypothesized to receive a lower score for the five attitude items with the exception of the impatience item which was hypothesized to get a higher score in the Extremely Bored condition (Table 6). The mood and personality conditions were compared with their opposite conditions, e.g., Positive Valence versus Negative Valence. To investigate these differences, Wilcoxon signed rank tests were conducted on each questionnaire item.

The extremely positive audience was perceived to be significantly more interested (z = -3.41, p = 0.001) in the talk, more positive (z = -2.25, p = 0.03) towards the talk, more eager to get information (z = -3.12, p = 0.002), and less impatient (z = -3.10, p = 0.002) than the extremely bored one. This finding completely matches with the parameter settings of the Extremely Positive and Extremely Bored conditions (Table 6). The Critical condition was perceived to be similar to the Extremely Positive condition, with one exception that the audience was significantly more critical (z = -2.18, p = 0.03) than the extremely bored audience while the extremely positive audience was not. This result also matches with the parameter settings of Critical condition, except for the Approval item. The Positive condition was also perceived to be similar to the Extremely Positive condition, but the positive audience was not significantly more positive or more eager to get information than the extremely bored audience was. This result is consistent with the parameter settings of the Positive condition except again for the Approval item.

To further investigate the attitude conditions, the items of Interest, Approval, Eagerness for Information, and Impatience in these conditions were compared with those in the Extremely Positive condition that were all set at high levels. The Criticism items in these conditions were compared with that in the Critical condition which was also set high. The bored audience was found to be significantly less eager (z = -2.13, p = 0.03) to get information than the extremely positive audience. This suggests that the Bored condition differentiated from the Neutral condition, which showed no significant difference from this high-level condition. Additionally, the extreme positive audience was found to be significantly less critical (z = -2.33, p = 0.02) than the critical audience.

For the mood and personality conditions, no significant difference was found to support the second and third hypotheses. Still, the audience with a positive valence parameter setting was perceived to be more aroused (z = -1.97, p = 0.049) than the audience with a negative valence parameter setting. Also the audience in Extremely Bored condition was rated as less extrovert than audience in the Critical (z = -2.54, p = 0.01), Extremely Positive (z = -2.28, p = 0.02), and Positive (z = -2.12, p = 0.03) conditions.

2.9. DISCUSSION AND CONCLUSIONS

This study has built an audience model that significantly predicts audience behavior using the agent parameters of mood, attitude, and personality. The audience model can generate expressive behavior by setting the model parameters. Both results of free-description and factor rating evaluation show that people can perceive variations in the attitude of the virtual audience that are caused by manipulation of the corresponding agent attitude parameters, which supports the first hypothesis. The results did not find similar matching between agent parameter manipulation and individual's perceptions of the virtual audience when it came to audience's mood (*H2*) and personality (*H3*). However, manipulation of the agent's valence parameter had an effect on perceived level of audience's arousal. This may be caused by a correlation between valence and arousal, suggested by many studies on the affective space, e.g., [38]. Furthermore, the virtual audience's expressiveness of moods might have been limited. Adding facial expressions might enhance this as various reports confirm that virtual characters express emotion better when using multi-modal expressions, e.g., [26]. Likewise, variation in audience's personality, in this case extraversion, was observed in the extremely bored, positive and critical attitude conditions. In other words, personality trait variation was only perceived in the attitude conditions that were more complex, i.e., created by multi-parameter manipulation. These conditions might enable more behavioral variations of the audience, which exhibited the personality traits. This is essential as the ability to express and perceive a person's individuality is situation dependent [39]. For example, someone's personality might be easier to assess when observed at a neighborhood party than as a soldier in a military parade. Furthermore, the expressiveness of the personality parameters could also have been constrained by the scope of the corpus. Although already extensive with 9600 coding units, the corpus was obtained by observing 16 individuals with personalities that did not cover the entire spectrum of personality trait combinations.

This study can be extended in many directions. First, control over the simulation environment can be added to easily construct different scenarios. For example, the classroom could be adjusted to a business meeting with fewer people sitting around a table or to a large podium with a larger audience, or the scenario can be changed from public speaking to musical performance. Second, the model could be extended by including social influence among individuals, as suggested by Poeschl and Doering [40]. Third, the functionality of the perception module in the model could be extended by adopting the Perception Markup Language (PML) [41] standard so that new perception technology by other researchers can be integrated. Fourth, the effect of changes in the settings of individuals' parameters on the output behavior and perceptions of this could be studied. Besides providing insight into which changes result in noticeable behavioral changes, this might also inform theories about audience behavior. Finally, potential operators (e.g., therapists) could be involved in the design of specific audiences to meet their needs.

Although the behavioral model for the autonomous agents was designed to create a virtual audience, a similar model might also be applied to VR systems that need autonomous virtual humans in other social situations. This could be for psychotherapy concerning disorders such as paranoia [42] and agoraphobia [43].

The evaluation presented in this paper focuses only on how people perceive the virtual audience behavior. This is an important validation step before claims can be made that a specific virtual audience setting (e.g., critical attitude) has a specific effect on users' emotional state in the future.

In conclusion, the main contributions of this work are as follows: (1) an audience model for public speaking simulation systems that generates expressive behavioral styles flexibly by adjusting agent parameters of mood, attitude, and personality, and (2) a corpus[1] of audience behavior showing different attitudes in public speaking situations and a coding scheme for posture observation. This audience model was built using a statisti-

[1] The corpus is available at http://dx.doi.org/10.4121/uuid:d613cc9c-c10b-4c50-be50-ba8ef7885dc5

cal approach based on observations of real audiences in public speaking situations. Using the parameters of attitude, mood, and personality as predictors, the audience model significantly predicted the audience behavior. This model was applied to an audience simulation, and the evaluation results showed that the virtual audience can behave expressively with regard to their attitude, and the behavioral styles can be controlled by modifying the model parameters. This is an important step towards providing users with a flexible and dynamic virtual environment in which they can be exposed to a virtual audience, for example, as part of a psychological stress test procedure, training, or psychotherapy.

REFERENCES

[1] C. Zanbaka, A. Ulinski, P. Goolkasian, and L.F. Hodges, "Social Responses to Virtual Humans: Implications for Future Interface Design," in *Proc. SIGCHI Conf. Human Factors in Computing Syst*, 2007, pp. 1561–1570.

[2] M. Slater, D.-P. Pertaub, C. Barker, and D.M. Clark, "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*, vol. 9, no. 5, pp. 627–33, Oct. 2006.

[3] J. Bissonnette, F. Dubé and M.D. Provencher, "The Effect of Virtual Training on Music Performance Anxiety," in *Int. Symp. Performance Science*, 2011, pp. 585–590.

[4] M. Powers and P.M.G. Emmelkamp, "Virtual Reality Exposure Therapy for Anxiety Disorders: A meta-analysis," *J. Anxiety Disorders*, vol. 22, no. 3, pp. 561–569, 2008.

[5] O. Kelly, K. Matheson, A. Martinez, Z. Merali, and H. Anisman, "Psychosocial Stress Evoked by a Virtual Audience: Relation to Neuroendocrine Activity," *CyberPsychology & Behavior*, vol. 10, no. 5, pp. 655–62, Oct. 2007.

[6] C. Kirschbaum, K.-M. Pirke, and D.H. Hellhammer, "The 'Trier Social Stress Test'-a Tool for Investigating Psychobialogical Stress Responses in a Laboratory Setting," *Neuropsychobiology*, vol. 28, no. 1–2, pp. 76–81, 1993.

[7] S.E. Taylor, T.E. Seeman, N.I. Eisenberger, T.A. Kozanian, A.N. Moore, and W.G. Moons, "Effects of a Supportive or an Unsupportive Audience on Biological and Psychological Responses to Stress," *J. Personality and Social Psychology*, vol. 98, no. 1, pp. 47–56, Jan. 2010.

[8] S.F. Hofmann and M.W. Otto, *Cognitive Behavioral Therapy for Social Anxiety Disorder: Evidence-Based and Disorder-Specific Treatment Techniques*, 1st ed. Routledge, 2008.

[9] P.M.G. Emmelkamp, "Behavior Therapy with Adults," in *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change*, 6th ed., no. 1, M.J. Lambert, Ed. Wiley, 2013.

[10] R.G. Heimberg and R.E. Becker, *Cognitive-Behavioral Group Therapy for Social Phobia: Basic Mechanisms and Clinical Strategies*. Guilford Publications, 2002, p. 334.

[11] M.J. Gallego, P.M.G. Emmelkamp, M. Van der Kooij, and H. Mees, "The Effects of a Dutch Version of an Internet-Based Treatment Program for Fear of Public Speaking: A Controlled Study," *Int. J. Clin. Health Psychol.*, vol. 11, no. 3, pp. 459–472, 2011.

[12] V. Vinayagamoorthy, A. Steed, and M. Slater, "Building Characters: Lessons Drawn from Virtual Environments," in *Proc. Toward Social Mechanisms of Android Science: A CogSci 2005 Workshop*, pp. 119–126.

[13] C.A. Pollard and J.G. Henderson, "Four Types of Social Phobia in a Community Samply," *J. Nerv. Ment. Dis.*, vol. 176, no. 7, pp. 440–445, 1988.

[14] J. Lee and S.C. Marsella, "Predicting Speaker Head Nods and the Effects of Affective Information," *IEEE Trans. Multimedia*, vol. 12, no. 6, pp. 552–562, 2010.

[15] L. Huang, L.-P. Morency, and J. Gratch, "Virtual Rapport 2.0," in *Proc. 10th Int. Conf. Intelligent Virtual Agents*, 2011, pp. 68–79.

[16] Z. Wang, J. Lee, and S. Marsella, "Towards More Comprehensive Listening Behavior: Beyond the Bobble Head," in *Proc. 10th Int. Conf. Intelligent Virtual Agents*, 2011, vol. 6895, pp. 216–227.

[17] C. Busso, Z. Deng, and M. Grimm, "Rigid Head Motion in Expressive Speech Animation: Analysis and Synthesis," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 15, no. 3, pp. 1075–1086, 2007.

[18] J. Allwood, "Bodily Communication Dimensions of Expression and Content," in *Multimodality in Language and Speech Systems*, B. Granstrom, D. House, and I. Karlsson, Eds. Kluwer Academic Publishers, 2002, pp. 7–26.

[19] E. Bevacqua, E. Sevin, S.J. Hyniewska, and C. Pelachaud, "A Listener Model: Introducing Personality Traits," *J. Multimodal User Interfaces*, vol. 6, no. 1–2, pp. 27–38, Apr. 2012.

[20] S.J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 2nd ed. New Jersey: Prentice Hall, 2003.

[21] C. Qu, W.-P. Brinkman, Y. Ling, P. Wiggers, and I. Heynderickx, "Human Perception of a Conversational Virtual Human: an Empirical Study on the Effect of Emotion and Culture," *Virtual Reality*, online first, Aug. 2013.

[22] C.N. Moridis and A. A. Economides, "Affective Learning: Empathetic Agents with Emotional Facial and Tone of Voice Expressions," *IEEE Trans. Affective Computing*, vol. 3, no. 3, pp. 260–272, 2012.

[23] N. ter Heijden and W.-P. Brinkman, "Design and Evaluation of a Virtual Reality Exposure Therapy System with Automatic Free Speech Interaction," *J. CyberTherapy and Rehabilitation*, vol. 4, no. 1, pp. 41–55, 2011.

[24] D.P. Pertaub, M. Slater, and C. Barker, "An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience," *Presence: Teleoperators & Virtual Environments*, vol. 11, no. 1, pp. 68–78, 2002.

[25] W.A. Villaume and G.D. Bodie, "Discovering the Listener Within Us: The Impact of Trait-Like Personality Variables and Communicator Styles on Preferences for Listening Style," *Int. J. Listening*, vol. 21, no. 2, pp. 102–123, Aug. 2007.

[26] A. Kleinsmith and N. Bianchi-Berthouze, "Affective Body Expression Perception and Recognition: A Survey," *IEEE Trans. Affective Computing*, vol. 4, no. 1, pp. 15–33, Jan. 2013.

[27] "Short form for the IPIP-NEO (International Personality Item Pool Representation of the NEO PI)." [Online]. Available: http://www.personal.psu.edu/j5j/IPIP/ ipip-neo120.htm.

[28] M. Bradley and P.J. Lang, "Measuring Emotion: the Self-Assessment Manikin and the Semantic Differential," *J. Beh. Ther. & Exp. Psychiat.*, vol. 25, no. I, pp. 49–59, 1994.

[29] H.M. Rosenfeld, "Measurement of Body Motion and Orientation," in *Handbook of Methods in Nonverbal Behavior Research*, K.R. Scherer and P. Ekman, Eds. Cambridge: Cambridge University Press, 1982, pp. 199–286.

[30] C. Robson, "Coding Sequences of Behavior," in *Real World Research*, 2nd ed., Blackwell Publishing.

[31] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, Apr. 1960.

[32] K. Krippendorff, *Content Analysis: an Introduction to its Methodology*. Beverly Hills: Sage Publications, 1980.

[33] H. Grillon, F. Riquier, B. Herbelin, and D. Thalmann, "Virtual Reality as a Therapeutic Tool in the Confines of Social Anxiety Disorder Treatment," *Int. J. Disability and Human Development*, vol. 5, no. 3, pp. 243–250, Jan. 2006.

[34] D. Hartanto, N. Kang, W.-P. Brinkman, I.L. Kampmann, N. Morina, P.M.G. Emmelkamp, and M.A. Neerincx, "Automatic Mechanisms for Measuring Subjective Unit of Discomfort," in *Annu. Review of Cybertherapy and Telemedicine*, B. K. Wiederhold and G. Riva, Eds. Interactive Media Institute and IOS Press, 2012, pp. 192–196.

[35] A.J. Cowell and K.M. Stanney, "Manipulation of Non-Verbal Interaction Style and Demographic Embodiment to Increase Anthropomorphic Computer Character Credibility," *Int. J. Human-Comput. Stud.*, vol. 62, no. 2, pp. 281–306, Feb. 2005.

[36] L.K. Kammrath, D.R. Ames, and A.A. Scholer, “Keeping Up Impressions: Inferential Rules for Impression Change Across the Big Five,” *J. Exp. Soc. Psychol.*, vol. 43, no. 3, pp. 450–457, May 2007.

[37] W.-P. Brinkman, D. Hartanto, N. Kang, D. de Vliegher, I.L. Kampmann, N. Morina, P.G.M. Emmelkamp, and M. Neerincx, “A Virtual Reality Dialogue System for the Treatment of Social Phobia,” in *Proc. CHI 2012: extended abstracts on Human Factors in Computing Syst.*, pp. 1099–1102.

[38] P.J. Lang, M.M. Bradley, and B.N. Cuthbert, “International Affective Picture System (IAPS): Technical Manual and Affective Ratings,” 1997.

[39] D. Pennington, *Essential Personality*. Arnold, 2003, p. 284.

[40] S. Poeschl and N. Doering, “Designing Virtual Audiences for Fear of Public Speaking Training - an Observation Study on Realistic Nonverbal Behavior,” in *Annu. Review of Cybertherapy and Telemedicine*, B. K. Wiederhold and G. Riva, Eds. Interactive Media Institute and IOS Press, 2012, pp. 218–222.

[41] S. Scherer, S. Marsella, G. Stratou, Y. Xu, F. Morbini, A. Egan, A.S. Rizzo, and L. Morency, “Perception Markup Language: Towards a Standardized Representation of Perceived Nonverbal Behaviors,” in *Proc. 12th Int. Conf. Intelligent Virtual Agents*, 2012, pp. 455–463.

[42] W.-P. Brinkman, W. Veling, E. Dorrestijn, G. Sandino, V. Vakili, and M. van der Gaag, “Using Virtual Reality to Study Paranoia in Individuals with and Without Psychosis,” *J. CyberTherapy and Rehabilitation*, vol. 4, no. 2, pp. 249–251, 2011.

[43] K. Meyerbröker, N. Morina, G. Kerkhof, and P.M.G. Emmelkamp, “Virtual Reality Exposure Treatment of Agoraphobia: a Comparison of Computer Automatic Virtual Environment and Head-Mounted Display,” *Stud. in Health Technology and Informatics*, vol. 16, pp. 51–56, 2011.

3

The Design of Virtual Audiences: Noticeable and Recognizable Behavioral Styles

Expressive virtual audiences are used in scientific research, psychotherapy, and training. To create an expressive virtual audience, developers need to know how specific audience behaviors are associated with certain characteristics of an audience, such as attitude, and how well people can recognize these characteristics. To examine this, four studies were conducted on a virtual audience and its behavioral models: (I) a perception study of a virtual audience showed that people (n = 24) could perceive changes in some of the mood, personality, and attitude parameters of the virtual audience; (II) a design experiment whereby individuals (n = 24) constructed 23 different audience scenarios indicated that the understanding of audience styles was consistent across individuals, and the clustering of similar settings of the virtual audience parameters revealed five distinct generic audience styles; (III) a perception validation study of these five audience styles showed that people (n = 100) could differentiate between some of the styles, and the audience's attentiveness was the most dominating audience characteristic that people perceived; (IV) the examination of the behavioral model of the virtual audience identified several typical audience behaviors for each style. We anticipate that future developers can use these findings to create distinct virtual audiences with recognizable behaviors.

This chapter has been published in Computers in Human Behavior **55**, 680 (2016).

3.1. INTRODUCTION

VIRTUAL audiences can elicit responses in humans similar to those that are elicited by real human audiences [1], [2]. This is used in scientific research (e.g., [3]), psychotherapy (e.g., [4]), and training (e.g., [5]), because virtual environments are easier to configure and control than the real world. While some applications aim for a neutral audience (e.g., [6]), others may benefit more from an expressive audience. For example, the treatment manuals of exposure therapy [7], [8] suggest controlling the audience attitude as an effective means of controlling anxiety in a public speaking scenario; studies on stress responses explored variations of stress tests using supportive and non-supportive audiences [3], [9]. As virtual audiences in public speaking scenario are becoming more widely used, e.g., as part of the Trier Social Stress Test (TSST) [10], and in exposure therapy for social anxiety disorder, an empirically validated expressive virtual audience appropriate for these applications is needed.

When individuals are exposed to a virtual environment and perform in front of a group of virtual humans, their belief, anxiety, and performance can be affected. For example, Wallergard et al. [6] suggested that virtual audiences as part of a stress test can indeed, like human audiences, induce stress. Aymerich-Franch, Kizilcec, and Bailenson [11] used a virtual audience to study the effects of self-representation on public speaking anxiety. When presenting in front of a virtual audience, the individuals could see in a virtual mirror their virtual reflection which was manipulated to be similar or dissimilar to themselves. Others [12–14] focused on giving people the experience of performing in front of an audience as part of exposure therapy for individuals with social anxiety disorder. This experience has also benefited non-clinical applications. For example, Bautista and Boone [15] let teachers be trained with virtual students to master their skills of content delivery and student management. Likewise, Bissonnette et al. [5] trained performance arts students, in this case, young musicians to overcome their performance anxiety by performing in front of a virtual audience. The information expressed by virtual audiences can be used for various purposes. For example, the virtual audience in a public speaking training system manifested different attitudes as feedback for the speech performance [16]. Supportive and non-supportive audiences have been used to evoke different levels of anxiety [3], [9]. Thus, the expressiveness of a virtual audience, i.e., what information a virtual audience can express and whether people can recognize the information, becomes a key question when designing virtual audiences.

As virtual audiences are made up of individual virtual humans, the first step in the development is the generation of individual virtual humans with believable behavior. Extensive work has been done in simulating such behavior. This work ranges from facial expression of emotion [17], head movement [18], to full body posture simulation [19], [20]. Besides emotions, Chollet et al. [19] and Hu, Walker, Neff, and Tree [21] demonstrated that attitude and even personality of an individual virtual human can effectively be expressed by body language. To make the virtual characters believable, dynamic behaviors, i.e., displaying sequences of behaviors instead of still images, are often required. These sequences can be pre-scripted [21], computed by crafted rules that specify which behavior should be generated in a certain context based on psychological knowledge and literature [22], or generated by statistical models that predict body postures based on observation [19].

Besides the behavior as individual virtual humans, audience members also respond to each other's behavior. Although work has been done on crowd behavior [23] such as path planning and interaction between individuals of pedestrians, Kang et al. [24] specifically had looked at the interaction behavior in an audience. According to their audience model, when an individual audience member is looking at an audience member in the neighborhood, the member in the in the neighborhood responds by looking back.

Among various public situations, public speaking is a common scenario occurring in everybody's life, e.g., delivering a business proposal, teaching in class, or giving a speech at a wedding. In public speaking situations, body language is a main channel of expression for audiences. Knowledge about this is therefore essential for developers to develop audiences that can be tailored for the need of users at run time. Currently, studies on the effects of virtual audiences often used three audience styles, described as positive, neutral and negative (e.g., [9], [25]). Their results showed the benefit and potential of varying audience styles. However, no explicit and unified descriptions or guidelines could be found for designing such virtual audiences. Therefore, it is still a challenge for future studies that needs either similar or different audience styles.

Limited research has been devoted to audience behavior in public speaking scenario. Poeschl and Doering [26] and Tudor et al [27] provided some guidelines for behavioral design of realistic virtual audiences. They observed the behavior of a typical audience in a lecture and explored the behavioral patterns such as frequency, duration, and postural sequence of certain behavior category, e.g. paying attention. Kang, Brinkman, Van Riemsdijk, and Neerincx [24] proposed a parameterized audience model to generate expressive audience behavior for public speaking scenarios. The generated behavior was controlled by model parameters that defined the audience members' moods, attitudes, and personalities. They showed that the simulated audience using this model could behave expressively with regard to the audience attitude, and that the behavioral styles can be controlled by modifying the model parameters. Still, it is currently unclear about how an audience behaves underlying an audience style, e.g., a positive audience or a bored audience, and let alone which mood, attitude, or personality trait is associated with a specific audience style.

To simulate audiences for a variety of public speaking scenarios, more understanding about audience style and the relation with individual audience member characteristics is needed. These could be scenarios such as business people listening to an investment proposal pitch, employees assembled to hear the management announcement of potential layoffs, or students attending a Friday afternoon lecture who are eager to leave. Audiences in these settings clearly behave differently. To simulate these audiences, a key question is how people differentiate between these audiences. Regardless of the narrative or the way people are dressed, are people able to recognize different audience styles in a similar way people are able to recognize different facial expression independent of the context, such as anger or sadness? And what are these audience styles?

To address these issues, the work presented in this paper uses an existing virtual audience environment [24] to address four questions: (1) what variations in audience characteristics, in particular, mood, personality, and attitude, result in perceivable variations in audience behavior? (2) What combination of individual audience members' characteristics do people use to design prescribed audience styles? (3) What audience styles do people recognize and (4) what are the typical audience postures and behaviors associated with specific audience styles? To answer these questions the paper first describes a paired comparison perception experiment, which is a classic psychophysical method that was used to determine peoples' sensitivity in noticing a specific quantitative difference in an audience characteristic, e.g., higher or lower arousal. After identifying which audience characteristic resulted in noticeable audience behavior differences, people were invited to use these characteristics to design audiences for a set of public speaking scenarios. Clustering the audience scenarios based on the similarity of the characteristic settings resulted in five audience styles. Videos of virtual audiences were made for each style, and people were invited to match audience style description to each video. The last step of the study was to examine the parameterized audience model and identify specific audience postures and behaviors that were characteristic for the behavioral styles.

3.2. VIRTUAL AUDIENCE MODEL AND SIMULATION

THE work in this paper revolves around a parameterized audience model (Figure 3.1) [24] that underlies an audience of virtual humans in a virtual environment. This is a probabilistic model abstracted from observation of real human audiences. Behaviors of real audiences were recorded when they were listening to presentations on a topic they were interested in, were critical about, found boring, and were neutral about. The audience corpus [28] consists of 9600 coding units with a sampling interval of two seconds, specifying head, gaze, arm, hand, torso, and leg positions. To obtain a parameterized model, additional data was also collected about the audience members' personality (extroversion, agreeableness, openness, neuroticism, and conscientiousness), attitude towards the topic (interest, approval, eagerness for information, criticism, and impatience), mood (valence, arousal, and dominance), and energy level. To generate audience behaviors, 59 unique postures representing 80% of the corpus were grouped into 15 posture categories based on the probabilities of postures that would follow the current one in the observation corpus. The parameter data was used to train a series of logistic regression prediction functions for these 15 posture categories. Once the category is determined, the final posture of the virtual human is determined by random selection according to transition matrix of postures of the specific posture group. The full body posture of the virtual audience was updated every two seconds. The audience model also includes event response, for example, turning head if a phone goes off, or looking back when another audience member is looking at the virtual human. Full details about the virtual audience and the parameterized audience model can be found in the work by Kang et al. [24]. Setting these parameters creates different audience styles. Kang et al [24] showed that people could recognize different audience attitudes and perceive different degrees of attitudes in the audience simulation. In an attempt to make the mood of a virtual human more recognizable, the audience model in this paper was extended with facial emotion expression using the facial expression tool [17], which was directly controlled by the mood status of the virtual human. The facial expressions were static unless the mood status changed.

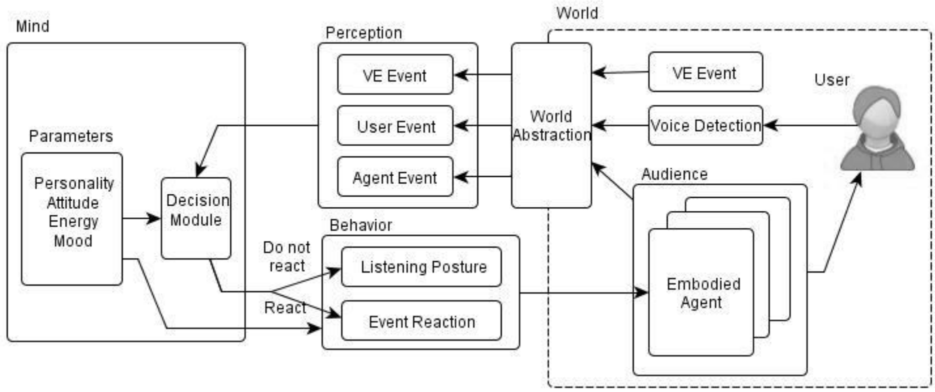


Figure 3.1: Framework of the virtual audience simulator.

3

3.3. STUDY I: PERCEPTION OF CHANGES IN PARAMETERS

3.3.1. RESEARCH QUESTIONS AND DESIGN

Kang et al [24] report that when people were asked to describe freely various audiences or rate their characteristics such as attitude, personality, and mood, no expected difference were found for manipulations of parameters such as extraversion, valence, or arousal. Although changes in these parameters led to observable changes of audience behavior, it was not clear whether people were actually able to recognize the changes in parameters. Thus, this study addresses the question what audience characteristics people can perceive. To answer this question, pairwise comparison, a classic psychophysical method, was therefore applied. This method provides more precise results in interval scales than a direct scaling, because it transforms the scaling task, which is difficult for humans, into a comparison task [29], [30]. For example, instead of being asked to rate directly the intensity of a specific characteristic of a virtual audience, participants are presented with two virtual audiences at the same time and asked which one they perceive to have a higher intensity of the characteristic.

3.3.2. EXPERIMENT SETTINGS

To determine perceivable audience characteristics, the first study investigated people's perception of the individual parameter changes. As some behaviors might only emerge when multiple parameters were changed, due to the correlation between the parameters in the observation corpus, the study also investigated people's perception when multiple parameters changed together as a group.

The parameter groups were determined by a principal component analysis on all the parameter data collected in the study by Kang et al [24]. The details of the analysis and grouping of the parameters are explained in Appendix A. The parameters with similar factor loadings, which indicated high correlations with each other, were grouped (Table 3.1), namely three independent parameter groups (IG1, IG2, and IG3), two independent single-parameter groups (IP1 and IP2), and three dependent parameters (DP1, DP2, and DP3). The independent groups correlated with different single factors, and the depen-

3

Table 3.1: Grouping of audience parameters and experiment settings

Parameters	Correlated factors	Grouping result	Number of levels for experiment	
			individual	group
Extraversion	1	IG1	3	3
Agreeableness	1		3	3
Conscientiousness	1		3	3
Openness	1, 4	DP1	3	3
Impatience	2	IG2	3	3
Eagerness for info	2		3	3
Interest	2		5	4
Approval	2, 3, 4	DP2	3	3
Dominance	3	IG3	3	3
Valence	3		7	4
Neuroticism	1, 3	DP3	3	3
Criticism	4	IP1	3	*
Arousal	5	IP2	5	*

*This parameter was not manipulated in a separate group condition.
Note: the loadings of the correlated factors are larger than 0.4.

dent parameters correlated with multiple factors. Therefore, the value of each parameter in the experiment was set by its correlated factors, i.e., when the factors were set, the values of the parameters were set.

The study investigated people's perceptions of the effects of 13 parameters in the model (Table 3.1) in two conditions: individual parameter adjustment and grouped parameter adjustment. Taking the perception of variation in the Interest parameter as an example here, in the individual parameter condition, only the Interest parameter was modulated, while in the grouped parameter condition three parameters of IG2 (i.e. impatience, eagerness for info, and interest) were modulated together. As the single-parameter groups (IP1 and IP2) only contained one parameter respectively, they were only included as individual parameters in the experiment. Additionally, in the grouped parameter condition, a combined question was used for independent parameter groups to see whether some parameters can be reduced to one control. For example, instead of only considering Interest, the participants were asked to give their overall opinion on the audience's Patience, Eagerness for information, and Interest together. For dependent parameters like Approval, no additional questions were asked. In both individual and group adjustment conditions, participants were asked to compare a few pairs of simulations in which the corresponding parameter or parameter group was set at different levels. Because participants' task load can be extremely high when the number of stimuli is large, the number of stimuli should be set as small as possible. To exclude unnecessary levels, the number of levels for each parameter or parameter group was determined in a pilot study. In the pilot study, the first author, who was regarded as the expert of this model, attempted 12 times to differentiate between pairs of simulations set at ten different levels, the maximum supported by the model. Accordingly, the maximum number of levels that the first author was able to recognize significantly ($p < 0.05$) was employed in the real perception experiment (Table 3.1). For the combined question in group adjustment condition, the number of levels was determined by the parameter in the group with the fewest number of levels, e.g., three levels for the Interest group (IG2).

3.3.3. Material and measures

An audience simulation was developed using the audience model mentioned in Section 2. To express mood better, facial expressions were added to this simulation using the facial expression tool [17]. Thus, the mood states of the audience affected not only the bodily responses, but also the facial expressions. To control the software load, only the audience members in the first row showed facial expressions in the audience simulation. An executable program was made for the pairwise comparisons, displaying two audience simulations side by side. The perception evaluation of parameters and parameter groups was conducted one by one separately, and the order was randomized. To evaluate the perceptions of one parameter (group) in N levels, N audience simulations were prepared with the parameter (group) varying from the lowest level (0) to the highest level (10) in the model. The program displayed sequentially one of all possible pairs (i.e., $N(N-1)/2$) of N audience simulations. The order of those pairs and the position (i.e. left or right) of the two audiences in each pair were randomly generated. For the evaluation of each parameter (group), a corresponding question was always displayed on the top, in the following form:

```
Which audience is more X?
```

According to the examined parameter (group), X was an adjective or a phrase, from the following list: pleased, aroused, dominant, open, conscientious, extroverted, agreeable, emotionally stable, positive towards the speech, interested in the talk, eager to get information, critical, and patient. All these terms corresponded to the audience parameters and were explained in an additional paper explanation card that included the definitions of the three dimensions of mood [31], the Big Five Factor of personality [32], and the audience attitude questionnaire (MA) used in the previous study [24]. For the combined question in the group condition, X was a combination of several terms which were involved in one group, e.g., “patient, interested, or eager to get information”for the Interest group.

3.3.4. Procedure

Hall [33] suggested females were better decoders of nonverbal behavior. Thus, the gender of participants was balanced so that gender difference in perception of nonverbal behavior could be examined. Twenty-four participants (12 females and 12 males) were recruited throughout the university campus to evaluate the audience model. Their age ranged from 24 to 41 years with a mean of 28.5 (SD =3.4) years. Each participant was asked to watch pairs of the audience simulations displayed on two desktop displays (iiyama ProLite E4315) respectively. Each pair of audience conditions displayed simultaneously for 20 seconds. After the simulations stopped, the participant was asked to answer the question which audience was more X, and then the next pair was displayed. When finishing all the comparisons for an individual parameter or a parameter group, the participant was asked about the rationale for the choices in the comparisons.

3.3.5. ANALYSIS AND RESULTS

The pairwise comparison data was analyzed using the method described in the study by Rajae-Joordens and Engel [30] (also described in Appendix B). The analysis was based on the Thurstone model [34], which provides scales on the differences people perceive among the stimuli. As post hoc analyses, multiple comparisons between simulation pairs were also conducted to examine whether these differences were significant. Confidence intervals of the differences, corrected by Scheffe's method, were used as the criterion for significance. Only when there were significant differences between the levels of a parameter, the adjustment of the parameter (group) was regarded as perceivable. Figure 3.2 shows the perceptual scales of all parameters and parameter goups. The scale was transformed that the value for the lowest level is always 0.0. These perceptual scale values show the relative locations of different levels of a parameter (group) on a psychological scale. That is, a value can be interpreted in terms of deviations from the values of other stimuli, and the deviations or differences follow a standard normal distribution. Taking Level 7 of valence in the individual condition for example (Figure 3.2a), it was about one standard deviation from Level 4 and about one and a half standard deviations from Level 2. The parameter levels were grouped by ovals according to the confidence interval test. Thus, the levels were perceived significantly different if they belong to the non-overlapping area of two different ovals on one scale. If some ovals are overlapping, the levels belonging to the overlapping part (e.g. Level 2 and 3 in Figure 3.2a, individual condition) do not show significant difference.

The results showed that people could differentiate between two levels for the perceivable parameters and parameter groups, with an exception of three levels for the perception of Valence. The mood states were well recognized in both individual adjustment and group adjustment conditions. For personality dimensions, agreeableness was recognized in individual adjustment condition, and neuroticism was recognized when adjusting the corresponding parameter group. The attitude items (i.e., Interest, Approval, Eagerness for information, and Impatience) were mostly perceivable in the group adjustment condition, while, in the individual adjustment condition, only Eagerness for information was perceivable. Thus, the audience behavior was more expressive in the group adjustment condition than in the individual adjustment condition for the attitude items.

To examine whether gender affects the perception, the Thurstone model of the perception for each parameter or parameter group was extended into a generalized linear model (GLM), taking both the perceptual scales and gender as the independent variables. The model fit was compared between the GLM and the Thurstone model (as described in the study by Rajae-Joordens and Engel [30]) respectively for each parameter or parameter group. Although males did not perceive the different levels of Approval as

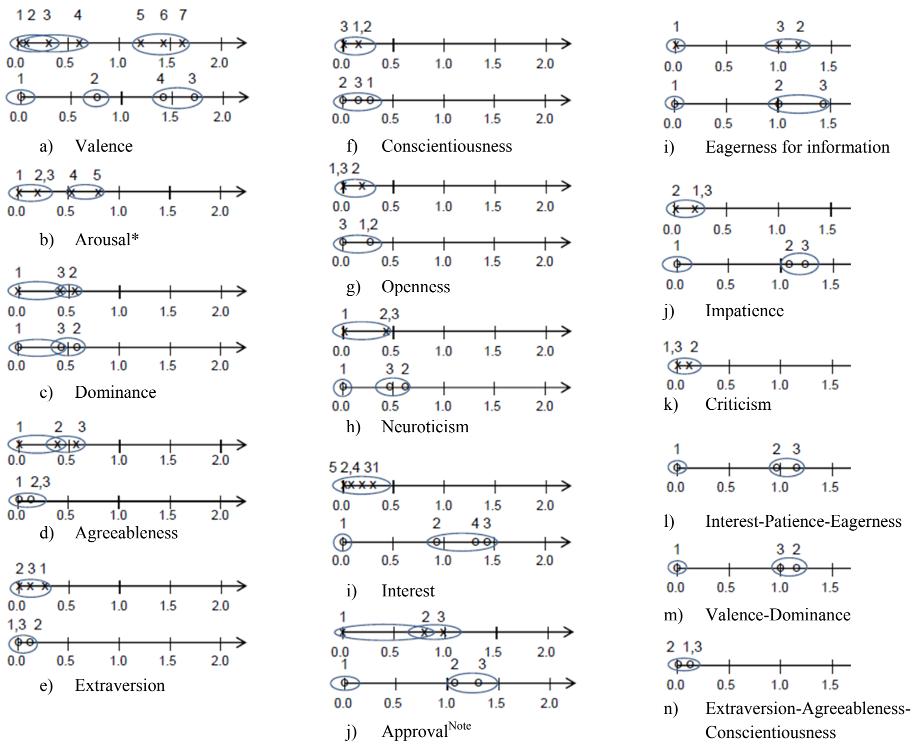


Figure 3.2: Perceptual scales of different levels of all parameters and parameter groups. The crosses (×) on the scales are the results from the individual adjustment conditions; the circles (∘) on the scales are the results from the group adjustment conditions. The levels are annotated by the numbers above the scales. The ovals demonstrate the statistical differences between the levels. Note: The figure for the group adjustment condition only includes the data of females instead of the whole sample (males and females).

the females did in the group adjustment condition (Figure 3.2j), no significant difference (p values ranging from 0.08 to 0.99, with a mean of 0.52, $SD = 0.29$) was found between the results of males and those of females for any parameter (group).

3.4. STUDY II: EXPRESSIVE AUDIENCE DESIGN

EXPRESSIVE virtual audiences are needed in various scenarios for different applications. Descriptions of audience behavior in these scenarios may provide direct information for the behavioral design. To this end, a design experiment was conducted to collect people's opinion on how audiences behave in different situations.

3.4.1. MATERIAL

A design interface (Figure 3.3) with one audience simulation and parameter controls was prepared so that participants could see the behavioral change when manipulating the audience parameters. The audience simulation model was similar to the one used in Study I. Unlike the study I where the virtual audience sat in a classroom, the audience for study II were seated in two rows in an arc with a blank background rendered in light blue. Without a specific background setting, the audience could fit into more scenarios. Controls of parameters were provided according to the results of the previous perception experiment (study I). Only the controls of parameters that were recognized in the previous study were provided so that the effect of parameter adjustment on the audience behavior was noticeable. In addition, participants could control parameters individually or as a group (i.e., some parameters increase or decrease together), according to the conditions in which the parameters were recognized. Thus, there were five controls for parameters which were recognized individually, namely, valence, arousal, dominance, eagerness for information, and agreeableness, and four controls for perceivable parameter groups, corresponded with Interest group (IG2), Valence-Dominance (IG3), Approval (DP2), and Neuroticism (DP3).

All the parameter controls allowed 3-level adjustment, i.e., low, medium, and high, and 0, 5, and 10 as the parameter value in the model, so that participant could always set medium level for the parameters. Although the perception experiment showed that people might only perceive two levels of some parameters, 3-level controls could avoid forcing participants to choose either high level or low level, thereby avoiding biased results. A reset button was also provided so that participants can reset all parameters to the medium level.

Twenty-three different audience scenarios (see appendix C) were described for participants to design the audience behaviors. Here are two examples of the scenario descriptions.

```
1) During a weekly school assembly for high school students,
the administrator is talking about the new rules the students
should obey.  The students find the rules much stricter than
before.

2) At a booth of an exhibition, an exhibitor is introducing a
new product.  People follow the explanation and find the design
innovative.
```

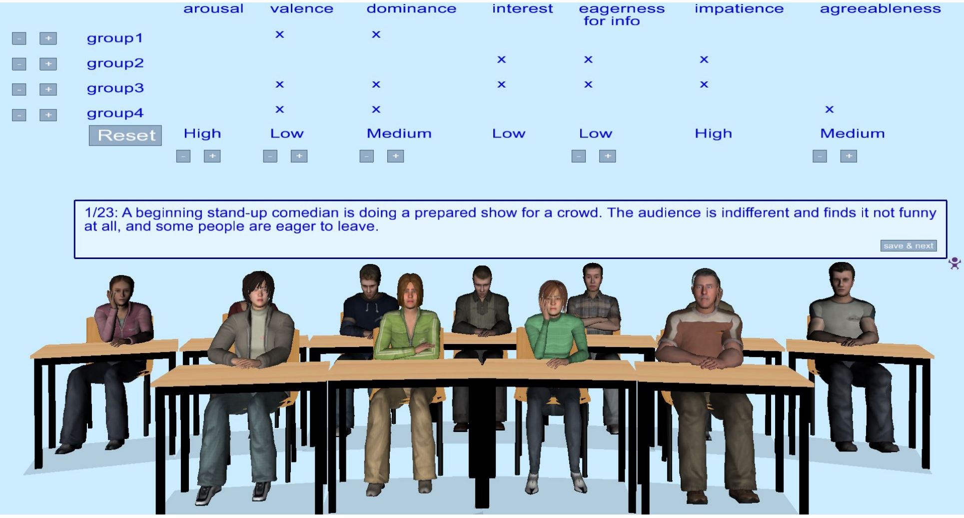


Figure 3.3: A screen shot of the design interface.

3.4.2. PROCEDURE

24 participants (12 females and 12 males) from 24 to 33 years old ($M = 27.5$, $SD = 2.5$) were recruited in the university. They were from seven different countries (14 Asians and 10 Europeans). They were asked to manipulate the audience behavioral styles by regulating some audience parameters so that the behavior matched the scenario descriptions. The paper explanation card about audience parameters was also given to the participants. The stimulated audience and the manipulation interface were shown on a TV (LG 42lm3450) about one meter from the position of participants. Before the experiment, participants practiced on manipulating the parameters to get some idea on how they could influence the behavioral styles. The order of scenario descriptions was randomized and participants were not allowed to go back to previous description and change the settings. After saving a setting, a new scenario description was given and all parameters were reset to the medium level.

3.4.3. ANALYSIS AND RESULTS

The first step of the analysis was to examine whether there was similarity in the way participants had designed the audience behavior for each scenario. To check this, the consistency of the settings for each scenario across the 24 participants was investigated. If the design of one scenario was inconsistent, it was expected that the adjustment options (i.e., high, medium, and low levels) for each parameter in this scenario were random, i.e., equally selected by the participants. Hence, the Chi-square tests of goodness-of-fit were performed for each parameter against an equal distribution of the three options to determine whether the three options were equally preferred. Besides the five individually controlled parameters mentioned in Section 3.1, the settings of Interest were also examined because Interest is affected by both Interest group and Approval so that its value can be different from that of Eagerness for information. The test results (Table 3.2) showed that most settings (122 out of 132) were significantly different ($p < 0.05$) from a random setting, and each scenario setting had at least three out of six parameters that were consistent across the participants.

The next step was to analyze how these obtained audience settings varied between scenarios and how these scenarios could be clustered. To do this, the similarity between each pair of scenarios was investigated. The setting of each scenario consisted of six parameters and therefore could be considered as a point in a six-dimensional space. The setting similarity between a pair of scenarios was calculated by taking the Euclidean distance between the two points representing the scenario settings in this six-dimensional space. That is, the shorter the distance was, the more similar the two settings were considered to each other. To examine whether or not the two settings were similar, the observed distance was compared with the expected distance, i.e., the average distance between two random points. See Appendix D on how the observed distance and the expected distance were calculated.

As each scenario was designed by 24 participants, there were 24 observed distances to examine the similarity between the setting of one scenario and the setting of another scenario. Thus, these 24 distances were compared with the expected distance between a pair of scenarios by a one-sample t-test. If the observed distances between a pair of scenarios were significantly ($p<0.05$) shorter than the expected distance, the two scenarios were regarded as similar to each other. According to the results, the scenarios were grouped into one category when they were similar to each other by showing significantly shorter distances between each other than the expected distance. The grouping results (Table 3.3) show that the scenario settings were grouped into five categories. The audience settings for scenarios 3 (the impractical business proposal) and 12 (the funeral eulogy) were excluded from the grouping because they did not show similarity in terms of short distance with any other scenarios.

The categorization was also inspected by applying a clustering method. An agglomerative hierarchical clustering method with complete linkage was employed to group the audience settings using Euclidean distances between all setting pairs as the measure. The idea was to build a tree of data that successively merges similar groups of settings. The similarity between each setting pair was measured using Euclidean distances between the two settings. According to the dendrogram obtained in the hierarchical clustering process (Figure 3.4), five categories were yielded by setting a distance threshold of 7.5 on a scale from 0 to 25, and Scenario 3 and 12 were excluded from the five categories. This categorization, if excluding Scenario 3 and 12, was exactly the same as that obtained by comparing the distances. Therefore, to obtain consistent setting features for each category, the audience settings were grouped into five categories, and Scenario 3 and 12 were excluded from the categorization.

The median settings of the five audience categories are shown in Table 3.4. The results provide an overview of the characteristics of each category. The difference in settings between the scenarios in different categories and similarity within a category suggest the potential existence of five distinct generic audience behavior styles.

3

Table 3.2: Results of chi-square tests on each parameter for each scenario, $\chi^2(2, N = 24)$

Scenario No.	χ^2 values for parameters					
	valence	arousal	dominance	interest	eager	agree
1	21.00	9.25	9.25	18.25	48.00	9.75
2	9.75	13.00	6.25	12.25	24.25	16.75
3	15.75	9.75	5.25*	7.00	3.25*	16.00
4	42.25	16.00	18.75	13.00	21.00	23.25
5	7.00	12.25	5.25*	10.75	24.25	14.25
6	14.25	10.75	1.75*	16.00	48.00	10.75
7	18.25	4.75*	6.75	13.00	32.25	12.25
8	18.25	19.75	18.75	18.25	13.00	9.00
9	12.25	24.25	24.25	14.25	48.00	12.25
10	42.25	28.00	12.25	12.00	16.00	15.75
11	21.00	19.75	0.75*	14.25	32.25	10.75
12	23.25	4.75*	5.25*	6.25	4.00*	12.25
13	14.25	12.25	3.25*	21.00	42.25	10.75
14	32.25	21.00	19.75	18.25	32.25	14.25
15	13.00	16.00	7.75	16.75	31.75	16.00
16	16.75	4.75*	1.00*	13.00	9.00	24.25
17	31.75	7.00	3.25*	12.00	21.00	14.25
18	21.00	12.25	21.00	18.25	37.00	7.75
19	19.00	4.00*	9.25	6.25	4.75*	5.25*
20	19.00	12.25	16.75	12.00	10.75	9.75
21	27.00	31.75	7.75	19.75	13.00	7.00
22	21.00	14.25	13.00	14.25	21.00	7.00
23	19.75	15.75	12.25	37.00	19.75	27.25

*The Chi-square value is less than 5.99, which is the critical value of $\chi^2(2)$ for $p = 0.05$. Thus, a result less than 5.99 indicates that the distribution was not found to deviate significantly from a random distribution.

3

Table 3.3: Grouping of the 21* audience scenario settings

Category	Scenario
A	1 (Promising business proposal), 4 (Best man's talk), 6 (Tuesday morning lecture about exam), 7 (Attractive Tuesday morning lecture), 9 (Related Monday morning meeting), 10 (Funny show), 14 (Positive corporate report), 18 (Innovative design), 20 (A qualified interviewee)
B	2 (Potential business proposal), 21 (A not very satisfactory interviewee), 22 (Training plan), 23 (Hobby talk)
C	16 (Announcement of stricter rules), 17 (Budget cut)
D	11 (Not funny show), 13 (Souvenir introduction), 15 (Repeated rule announcement)
E	5 (Friday afternoon lecture), 8 (Unrelated Monday morning meeting), 19 (Looking around in an exhibition)

*Scenario 3 (Impractical business proposal) and 12 (Funeral eulogy) were excluded.

Table 3.4: Median settings of the five audience categories

Parameter	Audience category				
	A	B	C	D	E
Valence	H	M	L	L	M
Arousal	H	M	H	L	L
Dominance	M	M	M	M	M
Interest	H	M	M	L	M
Eagerness	H	H	H	L	L
Agreeableness	H	M	L	L	M

L=0; M=5; H=10.

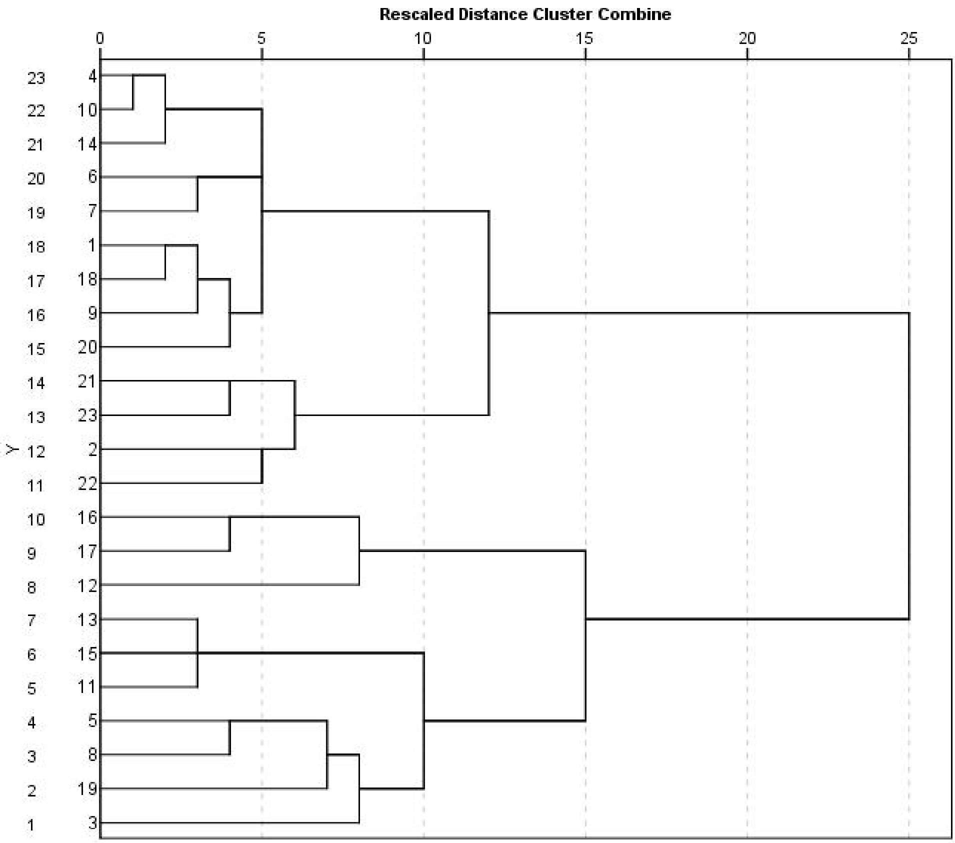


Figure 3.4: The dendrogram using complete linkage. The dash line specifies the threshold distance of 7.5.

3

3.5. STUDY III: PERCEPTION VALIDATION OF AUDIENCE SETTINGS

THE next step was to examine whether people were also able to recognize these five audience behavior categories. Thus, like other design-perception studies, (e.g., [35]), the design results need to be validated to ensure that people can recognize the designs. To validate the audience's behavior for different scenario descriptions, a perception experiment was conducted.

3.5.1. MATERIAL

The evaluation included the five settings (i.e., A, B, C, D, and E) found in the study II and a neutral setting as a baseline. To generate the different audience stimuli, the audience simulations from Study II were used. The five audiences were generated using the median settings (Table 3.4), and the neutral audience by setting all parameters to M. Three video clips of 30 seconds each were made for each audience setting from the audience simulation to avoid biased results, and two videos were randomly selected from the three and shown to participants. Thus, the evaluation consisted of 12 video clips: six settings and two clips for each. The 12 clips were displayed to participants in a random order to avoid the potential order effects.

3.5.2. MEASURES

A questionnaire was used to evaluate how well the virtual audience's behavior matched certain descriptions. The scenario descriptions were the same as those in the design experiment. The questionnaire was formulated as follows:

> `Which situations describe the audience in the movie best? Type A, B, C, D, or E? (Only one answer is possible.)`
>
> Type A: you may find such audiences in the following situations:
>
> [1] A person wants to start his own company and needs a sizable amount of investment money for this. He has an opportunity to introduce the investment proposal within 10 minutes to a number of business people, as they will consider whether or not they might invest in this new business opportunity. While listening, the investors find the proposal very promising.
>
> [2] The best man is talking about some interesting story about the new couple at a wedding party. The people in the party are mostly the new couples' family members and friends. They are friendly and enjoy very much the stories.
>
> [3] ...
>
> Type B, you may find such audiences in the following situations:
>
> ...

Table 3.5: Labels of different audience types

Audience type	Words and phrases collected	Label
A	^{2}Attentive, paying attention, ^{2}happy, ^{3}interested, ^{1}related to the audience, ^{3}enthusiastic, ^{1}engaged, ^{1}open to ideas, ^{1}easy, ^{1}positive,^{1} intrigued, ^{1}compliant, ^{1}pleased	Interested and enthusiastic
B	^{1}Hesitant, ^{2}slightly negative, ^{2}concerned, ^{1}topic is related to audience,^{1} uncertain, ^{2}cautious, ^{2}mixed opinion, ^{4}critical, ^{1}unclear	Critical and concerned
C	^{2}Anxious, ^{1}rebellious, ^{1}fearful, ^{1}not happy, ^{1}worried, ^{1}angry, ^{1}betrayed, ^{2}negative, ^{1}opposite, ^{1}tough, ^{1}threatened	Anxious and threatened
D	^{1}Inattentive, ^{1}indifferent, ^{4}bored, ^{2}impatient, ^{1}annoyed, ^{1}fed-up, ^{1}disengaged, ^{1}restless	Bored and impatient
E	^{1}Distracted, ^{1}impatient, ^{1}not really interested, ^{2}bored, ^{2}uninterested, ^{1}disengaged, ^{3}indifferent, ^{1}restless	Indifferent and uninterested

Note: the superscript number ahead of a word indicates that the number of people who mentioned the word, e.g., for type A, "interested" was mentioned by three people.

The audience situations for each type were explained by listing the full descriptions of scenarios (Appendix B) that were categorized as that type (Table 3.3).

As the description for each scenario type is very long, short labels might be more convenient for future use. Thus, several words and short phrases were collected to label the audience types from nine people (six females and three males, ranging from 25 to 36 years old, $M = 29.6$, $SD = 3.6$). Two words were selected to label each type, shown in Table 3.5.

To validate the labels, the participants were also asked to label the audience types by selecting one answer to the question as follows:

```
Which audience label describes the audience in the movie best?
(Only one answer possible)
```

To check whether the participants could read English and take the survey seriously, the following open question was also included about the participants' rationale for their choices.

```
Please fill in your rationale for your choices above in English.
Please provide specific answers, instead of a generic answer
for all videos.
```

Table 3.6: Agreement coefficients between audience descriptions and labels using different audience description scheme

Audience description scheme	A, B, C, D, E	A, B, C, Inattentive (D and E)	Attentive (A, B, and C), Inattentive (D and E)
Kappa	0.48*	0.65*	0.73*

*p <0.001.

3

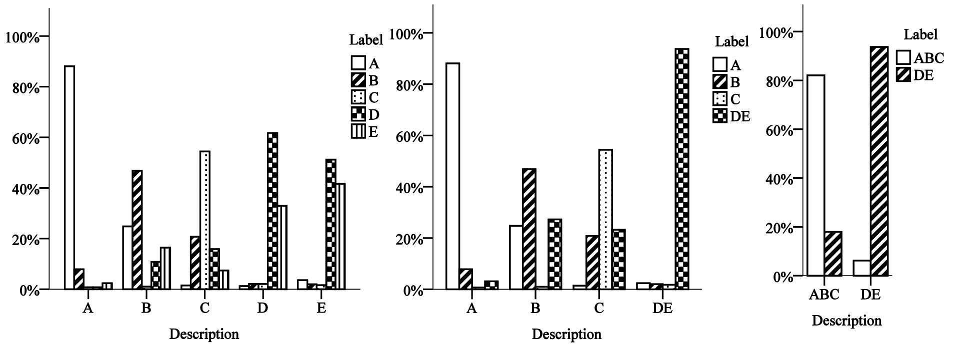
3.5.3. PROCEDURE

This survey was conducted online through Amazon Mechanical Turk among 101 people (51 females, 50 males) from the United States. Their age ranged from 18 to 64 years old (M = 34.9, SD = 10.9). The survey consisted of 12 parts. Each part contained one video clip and three questions; the participants were asked to watch the video and then answer the description and label questions, as well as an open question about their rationale. The video could be replayed. Once finishing one part and continue with the next part, participants were not allowed to go back to the previous part. At the start, participants were informed that they would only receive their one-dollar reimbursement if they follow all the instructions.

3.5.4. RESULTS

The result from one person was removed from the analysis because the answers to the open question were often obviously inconsistent with the choices of description and label. Before investigating how people perceived the videos, the descriptions and labels were examined to ensure that people could distinguish between the situation descriptions and between the audience labels. As the descriptions and the labels were expected to show one-to-one matches as designed, Cohen's kappa was calculated between the situation descriptions and audience labels. An agreement coefficient of 0.48 showed an overall significant (p <0.001) association between the labels and descriptions. Figure 3.5a shows the overview of the relationship between descriptions and labels. For descriptions A, B, C, and D, the most chosen label for each description was always the corresponding one, i.e., A, B, C, and D respectively. However, results did not always show one-to-one matches, e.g., descriptions D and E respectively matched with both labels D and E. Therefore, the descriptions and labels generally showed one-to-one matches, but the participants might not always be able to differentiate between the descriptions or labels of some types, e.g., type D and E. As people did not agree strongly on the five-type categorization, different audience description schemes were explored to obtain a more reliable description scheme. New description schemes were constructed by combining the types whose descriptions or labels might not be differentiated, e.g., type D and E. The coefficients (Table 3.6) increased to an acceptable level of 0.73 when audience types A, B, and C were integrated into one category and D and E into another, creating an attentive and an inattentive audience group. Thus, people showed more agreement on the two-type categorization. This suggested that most noticeable feature in the audience behavior was whether or not an audience paid attention to the presenter.

To investigate whether participants distinguished the different audience types, par-



a) five-type categorization b) four-type categorization c) two-type categorization

Figure 3.5: An overview of description-label relationship using different audience categorization schemes. The label choice distributions are expressed as a percentage for each description (category), i.e., the percentages of labels for one situation description (grouped bars filled with different patterns) should add up to 1.

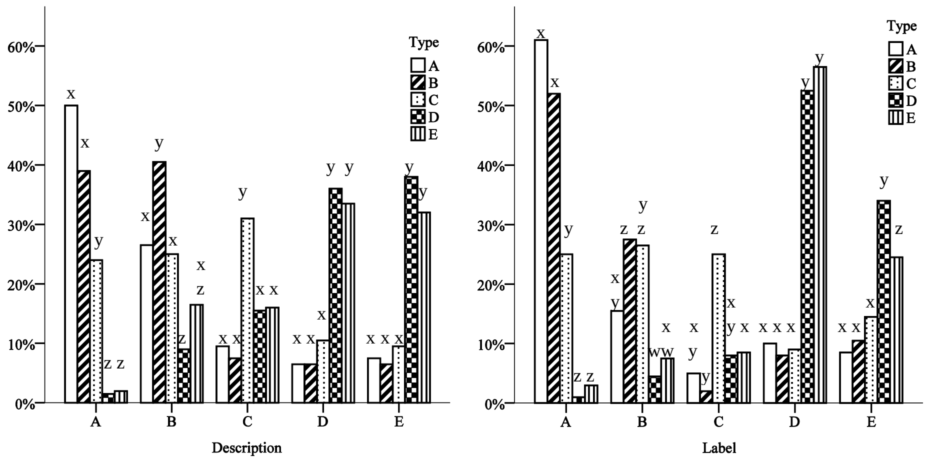
Table 3.7: Number of participants (n = 100) who categorized the audience as Inattentive

Audience	Sample 1		Sample 2	
type	Description	Label	Description	Label
A	17	23	11	14
B	12	16	14	21
C	22	29	18	18
D	75	87	73	86
E	63*	80	68	82

*p=0.12, the p value of all other results without notation is p <0.001.

ticipants' perception was tested across the audience types, using the binominal description scheme, i.e., attentive and inattentive. The numbers of participants who categorized the types as inattentive were tested by binomial tests against a random proportion of 50%. The test was conducted respectively on the descriptions and the labels for each sample video. As shown in Table 3.7, audiences A, B, and C were mostly (p <0.001) categorized as attentive while most (p <0.05) people categorized D and E as inattentive.

Besides the obvious difference between attentive audiences and inattentive audiences, the differences within each group were also investigated. Figure 3.6 shows an overview of the choice distribution of descriptions and labels for different audience types. Friedman tests were conducted on each description and label choice to find out whether the choices varied across the audience types. The results (Table 3.8) show an overall significant (p <0.001) difference in the choices of each description and each label among the audience types. Pairwise comparisons were further conducted using z-test on the proportions of each description or label between different types. The significance values of the z-tests had been adjusted using the Bonferroni method. The results are also shown in Figure 3.6. Description A and label A were most preferred by the participants for type A, although not significantly more than description B and label B; description



a) Participants' choices of descriptions b) Participants' choices of labels

Figure 3.6: An overview of participants' choices of descriptions and labels for different audience types. The description and label choice distributions are expressed as a percentage on each condition, i.e., the percentage of each description or label in one type (bars filled with the same pattern) should add up to 1. Each letter on top of a bar denotes a subset of audience type categories whose proportions for a certain description or label do not differ significantly (p <0.05) from each other.

Table 3.8: Results of Friedman tests of description and label choices across the audience types, $\chi^2(4, N = 200)$

	Audience type				
	A	B	C	D	E
Description	197.33	60.57	48.51	115.83	120.31
Label	275.23	62.63	71.87	239.21	56.57

Note: the p values of all statistics are less than 0.001.

B was significantly (p <0.05) more preferred for type B than other conditions while label B was preferred for both types B and C; both description and label C were selected by significantly (p <0.05) more participants in type C; no significant difference was found between types D and E, with similar preference for descriptions and labels D and E. Therefore, both description and label results showed the differences within the attentive audiences (A, B, and C), but no difference was found within inattentive audiences (D and E). However, some labels did not show a one-to-one match for a certain type. For example, label B was found often chosen for both types B and C, and label D were more chosen than label E by the participants to describe both types D and E.

In conclusion, the difference between attentive (A, B, and C) and inattentive (D and E) audiences was perceived to be significant. The descriptions and labels for types A, B, and C were partially validated: although people could not always perceive the difference between A, B, and C, there was a trend that the corresponding descriptions and labels were often the most preferred. No significant difference was found between types D and E.

3.6. STUDY IV: THE BEHAVIOR OF THE AUDIENCE TYPES

To gain some insight about people's perception of the audience types, the audience behavior was investigated for the different audience types. The study was conducted for each audience type on three aspects: (1) the specific postures, (2) the frequency of body movements, and (3) reactions towards disruptive events.

3.6.1. LISTENING POSTURES OF THE PERCEIVABLE AUDIENCE TYPES

To examine the specific behavior for the perceivable audience types, the statistical model for the generation of listening postures was used. The median settings of five audience types obtained in Study II were used as the model input, and thus one posture category was obtained for each audience type.

Table 3.9 summarizes the audience behavior for each type. Concerning the head position, a trend of decreasing attention was observed in order of type A, B, C, and D. While the head for type A was always facing the front, two out of three head positions for type D were looking downwards. A trend of decreasing openness showed in the position of arms and hands from type A to C. Compared with type A, type B audience showed more closed gestures, e.g., clenched hands and folded arms, and the gestures in C are totally closed. Type C also differed from A and B in the torso position. The upright position suggested less relaxation in the torso. Apart from the head position, type D also distinguished well from A, B, and C in arm and leg position, e.g., the fidgeting hands and legs. The behaviors for type D and E were almost the same except the facial expression. However, as the audience's heads were mostly lowered, it might explain that a difference was hardly recognized in Study III.

Table 3.9: Audience behavior in different settings

		A Interested and enthusiastic	B Critical and concerned	C Anxious and threatened	D Bored and impatient	E Indifferent and uninterested
Mood setting [Valence, Arousal, Dominance]		H, H, M 	M, M, M 	L, H, M 	L, L, M 	M, L, M 
Listening Posture	Number of postures	2	4	3	3	3
	Head	Upright	^{3}Upright; ^{1}tilted position, facing the front	^{1}Upright; ^{1}lowered head ^{1}tilted head	^{1}Upright; ^{2}lowered head	^{1}Upright; ^{2}lowered head
	Arms and hands	^{1}hands open on desk; ^{1}one hand touching or holding the other arm	^{1}clenched hands resting on desk; ^{1}hands open on desk; ^{1}one hand touching the neck, with the other resting on the front torso; ^{1}folded arms	^{2}arms folded; ^{1}hands on legs	^{2}suporting the head; ^{1}one or two hands tap on the desk continuously	^{2}suporting the head; ^{1}one or two hands on tap the desk continuously
	Torso	^{1}Torso forward; ^{1}Torso backward.	Torso backward	Torso upright	Torso forward	Torso forward
	Legs	Crossed or twisted legs.	Crossed or twisted legs.	Crossed or twisted legs.	^{2}both feet flat on the floor; ^{1}leg joggling or feet tapping	^{2}both feet flat on the floor; ^{1}leg joggling or feet tapping
Average probability of posture shifts		0.03	0.04	0.11	0.06	0.06
Reaction to an event no longer than 4 seconds		No	React	React	React	React
Reaction to an event no shorter than 5 seconds		React	React	React	React	React

Note: the superscript numbers indicate how many postures of the corresponding audience type comprise this position. If a position is not specified, all the postures of this audience type comprise this position.

3.6.2. Bodily movements

Another factor that affects the perception of audience behavior is the frequency of bodily movements. The bodily movements include posture shifts and consistently moving behaviors such as finger tapping. To study how often the behavior shifted from one posture to another, the posture transition matrix of the audience behavior model was inspected. The transition matrix presented the probabilities for each posture to transition to other postures in the successive time unit, i.e., two seconds in the model. Thus, the probabilities for posture shifts were used as one measure of bodily movement. To judge how the frequencies differed across the audience types, the average of the probabilities within one posture category for each audience type was calculated as the measure (also shown in Table 3.9). The results present an increase in the probability of posture shifts in the order of audience type A, B, and C. Although type D audience shifted their postures less often than type C, it actually exhibits much more bodily movements, because two out of three postures involve consistently moving behaviors such as finger tapping. As type D and E employed the same posture category, the movement probability was also the same. This suggested that an inattentive audience might exhibit more bodily movements than an attentive audience.

3

3.6.3. Event reaction

Besides listening behavior, the different audience settings may also affect the audience member reaction to disturbing events, specifically, whether or not to respond to such an event, such as a door slam. To study the reaction, the audience settings were used as the input of the reaction decision function [24]. As the reaction was also related to the event duration, the event duration was set 0.5, 1, 2, 3, 4, and 5 seconds respectively. The results (shown in Table 3.9) show that when the event duration was short, i.e., no longer than four seconds, only type A audience would not respond, while all other types would respond. However, if an event was long enough to be distracting, the most attentive audience would also respond.

3.7. Discussion and Conclusions

In conclusion, people were able to perceive changes in some of mood, personality, and attitude parameters by observing a virtual audience's behavior. Using the perceivable parameters, several audience scenarios were constructed by a group of individuals who acted as designers of virtual audiences. Audience parameter settings of individual audience scenarios showed extensive consistency across these designers. Furthermore, the list of 21 audience scenarios could be clustered into five underlying generic audience behavior styles. This led to the creation of five audience styles using the median settings of each cluster group. The perception validation study of these five styles showed a dominating characteristic of an audience that people perceived was whether or not the audience was attentive or inattentive. Although weaker, the findings also suggested that people could distinguish between interested-enthusiastic audience, critical-concerned audience and anxious-threatened audience. Finally, the findings of study IV gave an overview of the audience behaviors that made up these five audience styles. We anticipate that future developers can use these to create different recognizable audience styles.

As suggested in the previous study [24], facial expressions were added to improve the virtual audience's expressiveness. This point was supported by the results of Study I and Study II. However, according to the rationale provided by participants in the open question in Study III, it seemed that many participants did not use facial expressions as a clue to the judgment of audience styles. For example, only 18 out of 100 participants mentioned facial expression for videos of interested-enthusiastic (type A) audience. This might be caused by several reasons. First, the facial expressions in Study III were static in each condition. Without any change or comparison, people might not differentiate between some expressions, e.g., moderate pleasure and a neutral mood. Second, there was no control of the display devices the participants used to watch animations in the online survey. If the screen was small, participants might not have seen the facial expressions clearly as the scene included 11 characters in total. Nevertheless, Study III also showed that without any hardware constraints, people could recognize different audience styles. Besides facial expressions, another explanation of the lack of expressiveness in some situations is the underlying audience model used in the studies. As it was built based on observations of only 16 people in a university, the output behavior may lack variations for some audience types such as D and E.

This study can be extended in many directions. First, behavior of the current audience corpus was based on observations of normal conditions without extreme moods or attitudes. The behavioral model can be extended to show more extreme conditions by observing audiences in more diverse situations other than the classroom setting in the current corpus, e.g., business meeting and theatre. It is also worth exploring whether audiences' social-economic backgrounds or cultures influence their behavioral styles and whether people from more similar social-economic backgrounds or cultures would show more agreement about their perception of these audience styles. Second, to make a clear distinction between different audience styles, the parameter settings were always similar for all virtual audience members in the presented studies. However, such a homogeneous audience would hardly exist in real life. Studying a more heterogeneous audience would therefore provide more insights into behaviors and people's perception of more complex real life audiences. Third, in the studies there was no interaction between audience and the presenter. It would therefore be interesting to study a responsive virtual audience that would react according to the speaker's behavior. In this way, a presenter may perceive a stronger connection with the virtual audience, hence higher social presence [36]. Besides the perception of a virtual audience from bystanders' view as conducted in this paper, speaker's perception and responses could also be investigated, involving factors such as speech content, emotions, and the speaker's confidence.

When virtual audiences take the place of real audiences for various purposes such as psychotherapy and performance rehearsal, it is important that virtual audiences elicit similar responses in the users to those elicited by real audiences to ensure the effectiveness. For this, immersion, place illusion, and plausibility illusion are the three key concepts to understand [37]. Whereas immersion is a description of the characteristics of the system, e.g. the image quality of virtual audience, presence, i.e. place illusion, is related to the feeling of 'being there', and plausibility illusion refers to the illusion that the depicted scenario is actually occurring. A recent meta-analysis [38] on the relationship between presence and the intended provoked anxiety within virtual environments developed for psychotherapy of anxiety disorders, however, found no correlation between anxiety and presence experienced in virtual environments for treatment of social anxiety such as virtual audiences. Ling et al [38] pointed out that the presence measured in the studies mainly considered space illusion but not plausibility illusion, which might be a key issue when it came to social anxiety. In a similar manner, Poeschl and Doering [39] also stressed the need to understand people's experience of realism when exposure to scenario that involved a virtual audience. Future work therefore should focus on measuring the plausibility illusion when studying virtual audience, i.e. the illusion that the social interaction with the virtual audience is actually happening. For social situations such as public speaking, this illusion specifically relates to social presence which refers to users' perception on the virtual social characters and experience of their relationship with the virtual characters. Thus, social presence with virtual audiences could also be an important factor which affects users' responses to virtual audiences.

In conclusion, this paper explored audience simulation parameters, their settings and consequent audience styles, and validated them through a series of perception studies. This contribution is important because virtual audiences often function as key stimulus material. Validation is also vital as it provides the foundation for drawing any valid conclusions later on about people's behavior, emotions, and attitudes when they are exposed to these virtual audiences. The work also presents a more practical contribution by providing developers with guidelines for designing the behavior of virtual audiences. The potential existence of at least five underlying audience styles among the 21 public scenarios suggests that the five styles could represent a large variety of audiences which would occur in various public speaking scenarios. Thus, by implementing only five audience styles, designers would be able to construct many more different social settings with an audience, and users would have opportunities to experience more variations of social settings. As the parameter settings of the five audience styles also show consistency in the virtual audience's moods, attitudes, and personalities, designers should also consider the expressiveness of a virtual audience as a key factor to construct different audience styles successfully. Besides, as an audience's attentiveness is suggested as a dominating perceivable characteristic, it is an important characteristic to be mentioned and considered when describing or designing an audience. Additionally, the specific postures and behavioral patterns found in the five audience styles may help designers to develop virtual audiences with noticeable and recognizable behavioral styles. The findings can also be generalized to the design of individual virtual characters acting as listeners. Specifically, to design expressive virtual listeners, their behavior should variate in the following aspects: head and gaze direction, facial expression, frequency of bodily movements, reaction to disturbing events, and postural features such as openness, relaxation, and fidgets. The findings of the last study give designers directions on how to modulate these behaviors to create listening individuals as well as complete virtual audiences.

3

REFERENCES

[1] M. Slater, D.-P. Pertaub, C. Barker, and D. M. Clark, "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*, vol. 9, no. 5, pp. 627–33, Oct. 2006.

[2] C. Zanbaka, A. Ulinski, P. Goolkasian, and L. F. Hodges, "Social responses to virtual humans: Implications for future interface design," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2007, pp. 1561–1570.

[3] O. Kelly, K. Matheson, A. Martinez, Z. Merali, and H. Anisman, "Psychosocial stress evoked by a virtual audience: Relation to neuroendocrine activity," *CyberPsychology & Behavior*, vol. 10, no. 5, pp. 655–62, Oct. 2007.

[4] M. Powers and P. Emmelkamp, "Virtual reality exposure therapy for anxiety disorders: A meta-analysis," *Journal of Anxiety Disorders*, vol. 22, no. 3, pp. 561–569, 2008.

[5] J. Bissonnette, F. Dubé, M. D. Provencher, and M. T. Moreno Sala, "Virtual reality exposure training for musicians: Its effect on performance anxiety and quality.," *Medical problems of performing artists*, vol. 30, no. 3, pp. 169–77, Sep. 2015.

[6] M. Wallergård, P. Jönsson, G. Johansson, and B. Karlson, "A virtual reality version of the Trier Social Stress Test: A pilot study," *Presence: Teleoperators and Virtual Environments*, vol. 20, no. 4, pp. 325–336, Aug. 2011.

[7] R. G. Heimberg and R. E. Becker, *Cognitive-Behavioral Group Therapy for Social Phobia: Basic Mechanisms and Clinical Strategies*. Guilford Publications, 2002, p. 334.

[8] S. F. Hofmann and M. W. Otto, *Cognitive Behavioral Therapy for Social Anxiety Disorder: Evidence-Based and Disorder-Specific Treatment Techniques*, 1st ed. Routledge, 2008.

[9] S. E. Taylor, T. E. Seeman, N. I. Eisenberger, T. a Kozanian, A. N. Moore, and W. G. Moons, "Effects of a supportive or an unsupportive audience on biological and psychological responses to stress.," *Journal of personality and social psychology*, vol. 98, no. 1, pp. 47–56, Jan. 2010.

[10] C. Kirschbaum, K.-M. Pirke, and D. H. Hellhammer, "The 'Trier Social Stress Test'-A tool for investigating psychobialogical stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, no. 1–2, pp. 76–81, 1993.

[11] L. Aymerich-Franch, R. F. Kizilcec, and J. N. Bailenson, "The relationship between virtual self similarity and social anxiety," *Frontiers in Human Neuroscience*, vol. 8, p. 944, Nov. 2014.

[12] P. L. Anderson, M. Price, S. M. Edwards, M. a Obasaju, S. K. Schmertz, E. Zimand, and M. R. Calamaras, "Virtual reality exposure therapy for social anxiety disorder: a randomized controlled trial.," *Journal of consulting and clinical psychology*, vol. 81, no. 5, pp. 751–60, Oct. 2013.

[13] D. Hartanto, W.-P. Brinkman, I. L. Kampmann, N. Morina, P. G. M. Emmelkamp, and M. A. Neerincx, "Design and implementation of home-based virtual reality exposure therapy system with a virtual eCoach," in *Proceeding of the 15th International Conference on Intelligent Virtual Agents*, 2015, vol. 9238, pp. 287–291.

[14] N. Morina, W.-P. Brinkman, D. Hartanto, I. L. Kampmann, and P. M. G. Emmelkamp, "Social interactions in virtual reality exposure therapy: A proof-of-concept pilot study.," *Technology and health care: official journal of the European Society for Engineering and Medicine*, vol. 23, no. 5, pp. 581–9, Sep. 2015.

[15] N. U. Bautista and W. J. Boone, "Exploring the impact of TeachME™ Lab virtual classroom teaching simulation on early childhood education majors' self-efficacy beliefs," *Journal of Science Teacher Education*, vol. 26, no. 3, pp. 237–262, Feb. 2015.

[16] M. Chollet, G. Sratou, and A. Shapiro, "An interactive virtual audience platform for public speaking training," in *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, 2014, pp. 1657–1658.

[17] J. Broekens, C. Qu, and W. Brinkman, "Dynamic facial expression of emotion made easy," 2012.

[18] Z. Wang, J. Lee, and S. Marsella, "Towards more comprehensive listening behavior: Beyond the bobble head," in *Proceedings of the 11th international conference on Intelligent Virtual Agents*, 2011, vol. 6895, pp. 216–227.

[19] M. Chollet, M. Ochs, and C. Pelachaud, "From non-verbal signals sequence mining to Bayesian networks for interpersonal attitudes expression," in *Proceeding of the 14th International Conference on Intelligent Virtual Agents*, 2014, pp. 120–133.

[20] Y. Xu, C. Pelachaud, and S. Marsella, "Compound gesture generation: A model based on ideational units," in *Proceeding of the 14th International Conference on Intelligent Virtual Agents*, 2014, pp. 477–491.

[21] C. Hu, M. A. Walker, M. Neff, and J. E. F. Tree, "Storytelling agents with personality and adaptivity," in *Proceeding of the 15th International Conference on Intelligent Virtual Agents*, 2015, vol. 9238, pp. 181–193.

[22] E. Bevacqua, E. Sevin, S. J. Hyniewska, and C. Pelachaud, "A listener model: Introducing personality traits," *Journal on Multimodal User Interfaces*, vol. 6, no. 1–2, pp. 27–38, Apr. 2012.

[23] D. Thalmann and S. R. Musse, *Crowd Simulation*. London: Springer London, 2013.

3

[24] N. Kang, W. P. Brinkman, M. B. van Riemsdijk, and M. A. Neerincx, "An expressive virtual audience with flexible behavioral styles," *IEEE Transactions on Affective Computing*, vol. 4, no. 4, pp. 326–340, Oct. 2013.

[25] D. P. Pertaub, M. Slater, and C. Barker, "An experiment on public speaking anxiety in response to three different types of virtual audience," *Presence: Teleoperators & Virtual Environments*, vol. 11, no. 1, pp. 68–78, 2002.

[26] S. Poeschl and N. Doering, "Designing virtual audiences for fear of public speaking training - An observation study on realistic nonverbal behavior," in *Annual Review of Cybertherapy and Telemedicine*, B. K. Wiederhold and G. Riva, Eds. Interactive Media Institute and IOS Press, 2012, pp. 218–222.

[27] A.-D. Tudor, S. Poeschl, and N. Doering, "What do audiencences do when they sit and listen?," in *Annual Review of Cybertherapy and Telemedicine*, 2013, pp. 120–124.

[28] N. Kang, "Posture corpus for the behavior generator [Dataset]," 2013. doi:10.4121/uuid:d613cc9c-c10b-4c50-be50-ba8ef7885dc5.

[29] P. Engeldrum, *Psychometric Scaling: A Toolkit for Imaging Systems Development.* Imcotek Press, 2000.

[30] R. Rajae-Joordens and J. Engel, "Paired comparisons in visual perception studies using small sample sizes," *Displays*, vol. 26, no. 1, pp. 1–7, 2005.

[31] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament," *Current Psychology*, vol. 14, no. 4, pp. 261–292, Dec. 1996.

[32] L. A. Pervin, D. Cervone, and O. P. John, *Personality: Theory and research*, 9th ed. John Wiley & Sons, 2005.

[33] J. A. Hall, *Nonverbal sex differences: communication accuracy and expressive style.* Johns Hopkins University Press, 1984, p. 207.

[34] L. L. Thurstone, "A law of comparative judgement," *Psychological Review*, vol. 34, pp. 273–286, 1927.

[35] J. Xu, J. Broekens, and K. Hindriks, "Bodily mood expression: Recognize moods from functional behaviors of humanoid robots," in *proceedings of 5th international conference on social robots (ICSR)*, 2013, pp. 511–520.

[36] F. Biocca and C. Harms, "Defining and measuring social presence: Contribution to the networked minds theory and measure," in *Proceedings of the 5th International Workshop on Presence*, 2002, vol. 2002, no. 517, pp. 1–36.

[37] M. Slater, "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*, vol. 364, no. 1535, pp. 3549–57, Dec. 2009.

[38] Y. Ling, H. T. Nefs, N. Morina, I. Heynderickx, and W. P. Brinkman, "A meta-analysis on the relationship between self-reported presence and anxiety in virtual reality exposure therapy for anxiety disorders.," *PloS one*, vol. 9, no. 5, p. e96144, Jan. 2014.

[39] S. Poeschl and N. Doering, "The German VR Simulation Realism Scale--psychometric construction for virtual reality applications with virtual humans.," *Studies in health technology and informatics*, vol. 191, pp. 33–7, Jan. 2013.

4

Public speaking training in front of an imaginary or virtual audience: A randomized controlled trial

When preparing for a public speech, practicing with an audience is suggested to be effective in enhancing speech performance. However, it is often impractical to organize an audience to practice a presentation. Virtual reality can provide a solution, i.e., practicing with a virtual audience. This paper studied this practicing technique for enhancing speech performance and people's training satisfaction. A randomized controlled trial (n = 40) was conducted to compare practicing in front of a virtual audience with another practicing technique whereby the presenter had to imagine an audience while practicing. Individuals practiced their presentations in three training sessions with either a virtual audience or an imaginary audience. Participants' performance was assessed in a closing session where they delivered their speech in front of a human audience. The results showed that individuals seemed to benefit more from a virtual audience than an imaginary audience in reducing speech anxiety. The clearest benefit of practicing with a virtual audience was the satisfaction it gave. Participants were more positive towards training with a virtual audience regarding both the training process and its effect on their presentation ability. We anticipate that virtual audiences can be beneficial in motivating individuals to practice their presentation.

4.1. INTRODUCTION

WHETHER we are talking about our research at a conference, making a speech at a friend's wedding, or making a proposal in a business meeting, we have to speak in public from time to time. How well we have delivered a presentation affects the way people think about us and the message of the presentation. For example, people tend to regard presenters as more credible and intellectual when they have more eye contact with their audience during presentations [1], [2], and people are more likely to believe the presenters and be persuaded when the speech is fluent and well organized [3]. People's attitudes change more in the direction of message recommendations by strong arguments than weak arguments [4].

To deliver a well-received speech, preparation is necessary. Among all the speech preparation activities, Menzel and Carrell [5] specifically found that amount of rehearsals in front of an audience and the amount of experienced anxiety was associated with quality of the speech performance, i.e., how an individual deliver a speech in respect of body language, voice, confidence etc. Although the correlation between the amount of rehearsals and speech performance was not confirmed in a study by Smith and Frymier [6], they found that students who had practiced with an audience performed better than those who had practiced without an audience. Moreover, a study by Ayres et al [7] indicated that practicing with an audience helps to decrease public speaking anxiety and increase the willingness to speak.

Although practicing with an audience is suggested to be effective in enhancing speech performance, it is often impractical to organize an audience to practice a presentation. This problem can be solved by virtual reality, which can provide presenters with a virtual audience. The use of virtual audiences has already been suggested for several application domains. For example, they have been proposed to replace a human audience in the Trier Social Stress Test (TSST) [8] to induce stress in an individual with the aim of studying stress responses [9]. Speaking in front of an audience is also on the list of anxiety-evoking social situations [10]. Practicing in front of virtual audiences as part of a therapy for social anxiety is therefore also often studied [11]. Additionally, virtual audiences have been used to improve teachers' performance [12], [13].

To the best of our knowledge, little research has been reported on the use of virtual audiences in public speaking training. The work reported in this paper addresses whether practicing with a virtual audience could benefit the public speakers. It sets out to study this practicing technique for enhancing public speaking performance, focusing especially on its effectiveness and people's training satisfaction. The empirical study compared practicing in front of a virtual audience with another practicing technique whereby the presenter had to imagine an audience while practicing.

4.2. RELATED WORK AND HYPOTHESIS

Interaction with a single or a group of virtual characters can affect people in a manner similar to that of interaction with real people. For example, people have reported anxiety when speaking to a virtual character [14], [15], speaking among a group of virtual characters [16], or giving a presentation in front of a virtual audience [17]. The presence of a virtual audience has also been found to affect people's performance as a real audience did [18].

Without a visible audience, imagination can also affect people's emotion and performance. For example, imaginary audiences and virtual audiences were found similarly effective in evoking social anxiety [9]. Both practice methods may help to reduce people's anxiety during presentation and improve public speaking performance. Virtual audiences and imaginary audiences have therefore been used to treat social anxiety disorder, and were found comparatively effective in reducing social anxiety [19]. Furthermore, similarity in treatment effect for the phobic patients being exposed to a virtual or real audience has also been reported [20]. Many researchers also found positive correlations between rehearsal in front of audiences and performance [5], [6]. Moreover, some studies [21], [22] reported that imagination of a performance, or mental rehearsal of the process, can help to improve the performance. Success in performance can also boost self-efficacy beliefs [23]. For example, a system for training teaching skills [13] was found effective in improving teachers' performance and self-efficacy in teaching by providing a classroom simulation for teachers to practice with. Likewise, people's self-efficacy was influenced by comments made by virtual classmates about answers they gave in the class [16].

Still, practicing with an imaginary audience holds a number of drawbacks when comparing practicing with a virtual audience, which potentially makes it less satisfying, efficient and effective for the trainees to use. First of all, the imagery task itself has some limitations [24]. Imagining requires a great deal of attentional resources and most individuals have problems maintaining concentrated on the imagery task for a long period. In addition, the imagery is difficult to control because individuals may distract themselves from an undesirable image by replacing it with another thought or image which may weaken the effectiveness of such activities. Additionally, individual's capability of imagination may also affect the effectiveness. For example, individuals with no experience of speaking in front of a large audience may have difficulty in imagining such a situation. Furthermore, when practicing a presentation with an imaginary audience, it requires individuals to perform two tasks simultaneously: presenting and imagining. However, as dual-task performing requires more attentional resources, the problem of capacity overload may arise, which may deteriorate the performance of both tasks [25], [26]. Considering the limitations of the imagery task and the relative ease of practicing with a virtual audience, the following hypotheses are formulated:

> `People who practiced in front of a virtual audience (H1) perform better, (H2) are less anxious, (H3) find the practice method more satisfying, and (H4) hold more positive beliefs of their self-efficacy than those who practiced with an imaginary audience.`

4.3. METHOD

A public speaking training course was organized on the university campus. In this individual course, participants were instructed to practice their own presentations in either one of the two ways: practice in front of a virtual audience (i.e., VR condition), or practice with an imaginary audience (i.e., IM condition). To compare the effects of the practicing methods, a between-subjects design was employed.

4.3.1. MATERIAL

In the VR condition, participants could select out of four different virtual environments to match the presentation setting they were targeting. They were a classroom with normal seat arrangement, a classroom with U-shaped seat arrangement, a conference hall, and a setting for a public PhD defence (Figure 4.1). A parameterized expressive audience behaviour generator [27] was used to create an expressive virtual audience. The behaviour generator was accomplished based on statistical models abstracted from observation of real audiences. The audience corpus [28] obtained in the observation consisted of 9600 coding units with a sampling interval of two seconds, specifying head, gaze, arm, hand, torso, and leg positions. As shown in the audience framework (Figure 4.2), the generated behaviour was modulated by the audience members' personality (extroversion, agreeableness, openness, neuroticism, and conscientiousness), attitude towards the topic (interest, approval, eagerness for information, criticism, and impatience), mood (valence, arousal, and dominance), and energy level, which were also collected in the observation. The virtual audience showed different attitudes by their postures and facial expressions. In this study, participants could select an audience type, namely a positive, neutral, critical, or bored looking audience, to practice coping with the audience. Besides manifesting different listening postures, the virtual audience also respond to interrupting events within the virtual environment, including VE events such as loud noises and telephone rings and agent events such as being looked at by another audience member. For example, a bored audience always responded to such events by turning their heads whereas an interested audience only responded to events which lasted long. The perception of the four audience types have been validated in a previous user study [29]. Besides selecting an audience type, participants could also select whether an interrupting event would occur during the rehearsal of their presentation such as drilling noises or mobile phone rings. After the participants rehearsed the delivery of their talk, members of the audience asked the participants six questions, such as "What motivated you to carry out this research?" or "What would have improved your work?". These questions were randomly selected from a list of 40 questions respectively related to a specific scenario setting, e.g., academic conference or public thesis defence, which the participants had selected in advance. The questions were triggered by the use of voice detection. Three seconds after a participant stopped answering, the audience would pose the next question. If the participant did not answer, the audience member repeated the question.

Besides rehearsing the presentation, the course also included the use of a virtual coaching system both in VR and in IM condition. Using the system, a virtual coach provided information about presentation structure, body language during speech, and visual aids. The coach also helped participants to reflect on and improve their presentations. For this, participants entered their reflections into textboxes of the system to answer specific questions from the coach, which were designed to help participants to think about ways to prepare and improve their presentations, e.g., "How was your presentation going?" or "What can you do to improve your introduction part?". These answers were printed out for participants so they could use them to improve their presentation at home.

(a) Classroom with U-shaped seat positions

(b) Classroom with normal seat positions

(c) Conference

(d) Public thesis defence

Figure 4.1: Screenshots of different scenarios.

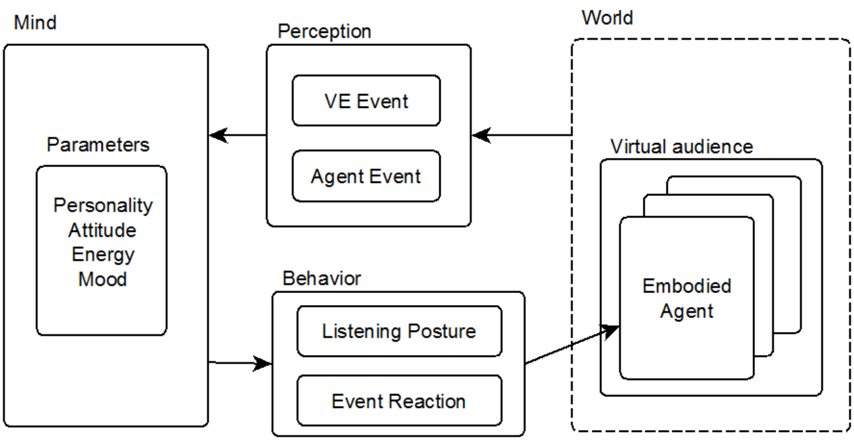


Figure 4.2: Framework of the virtual audience simulator.

4.3.2. MEASURES

OUTCOME MEASURES

Self-efficacy (SE). A one-item self-efficacy assessment was applied to measure self-efficacy in public speaking following the suggestions of Bandura [30]. The question was formulated as: *Supposing that now you need to give the presentation you are preparing for in the real situation, please rate how certain you are that you can successfully give the presentation.* The item was rated on a scale from 0 (highly certain cannot do) to 10 (highly certain can do).

Presentation Performance (PP). The rubric for oral presentations developed by the University of Wisconsin–Madison [31] was used to evaluate people's speech performance and presentation content after the training. The rubric consisted of ten items, which were rated on a four-point scale from 1 (poor) to 4 (excellent). The rubric consisted of four items evaluating the presentation content (i.e., subject knowledge, organization, visuals, and mechanics), five delivery-related items evaluating both nonverbal (eye contact, body language, and poise) and verbal skills (enthusiasm and speaking skills), and one timing item. In this study, the timing item was excluded as the presentation lengths and their timing requirement varied between participants.

Length of answers (LA). As behavioural assessment for social anxiety [32], [33], the lengths of the presentation and answers to the questions was taken as a measure for confidence or avoidance behaviour in the question phase of the closing presentation.

Utility questionnaire (UQ). To investigate how satisfying and useful people found the practicing methods, a 12-item utility questionnaire was designed consisting of five items evaluating the practice process and seven items on the effectiveness in improving presentation performance (Table 4.1). All the items were rated on a seven-point scale from 1 (strongly disagree) to 7 (strongly agree).

PROCESS MEASURES

Heart rate (HR). Heart rate is a physiological measure of experienced anxiety of people [34]. To measure the elicited fear responses during presentations, participants' heart rate was monitored continuously in each session. Besides, heart rate variability (HRV) was also analysed, which was found associated with mental workload and emotional strain [35], [36]. Subjective unit of discomfort (SUD) [37]. This item measures the levels of anxiety experienced by the participants. It is rated on a scale from 0 (no anxiety at all) to 10 (the highest level of anxiety that you can imagine).

Presence response (PR). Recently standard presence questionnaires have been criticized for their ability of measuring social presence [38]. Therefore, instead of measuring general feeling of presence, a three-item questionnaire on presence response was developed with the focus on how well realistic responses of people were elicited when practicing presentations. This questionnaire was adapted from the one used in a study by Pan et al. [39] (Table 4.1). It reflects presence by comparing participants' responses to what these would have been in a similar real situation – with respect to their overall behaviour, their emotional responses and their thoughts. Thus, compared with the immersive tendencies questionnaire (ITQ) [40] which measures individuals' tendency to be involved in virtual environments, this measure reflects more directly the practice process and how well the process can benefit the public speakers. All the items were rated

Table 4.1: Questionnaire items for measuring presence response and utility

Questionnaire and dimension		Item	
Presence response		Q1	To what extent did you find your emotional response and thoughts during the experience similar to those in a real presentation situation?
		Q2	How often did you find yourself almost automatically behaving within the room as if it was a real presentation?
		Q3	To what extent did you feel as if you were in a real presentation situation?
Utility questionnaire	Practice process	Q1	I like practicing in this way.
		Q2	I find such practice enjoyable.
		Q3	Little additional effort is needed to practice in this way.
		Q4	This method motivates me to practice.
		Q5	It is an efficient way to prepare my presentation.
	Presentation performance	Q6	It helps me to get used to speaking in front of an audience.
		Q7	It helps me to cope with different audiences.
		Q8	It helps me to become less nervous.
		Q9	It helps me to have more eye contact with the audience.
		Q10	It helps me to consider the audience.
		Q11	It helps me to consider unexpected events during presentation.
		Q12	It helps me to consider what questions I may get from the audience.

on a seven-point scale from 1 (not at all) to 7 (very much).

COVARIATES

Personal report of confidence as a speaker (PRCS) [41]. It is a 30-item self-report scale, which assesses both behavioural and affective responses to public speaking situations. The questions are answered in a true–false format, and the questionnaire score ranges from 0 (i.e., no fear of public speaking) to 30 (i.e., highest level of fear). This measure was used to investigate whether individual characteristic was associated with the effect of training.

4.3.3. PROCEDURE AND APPARATUS

Forty-eight participants (16 females and 32 males) were recruited throughout the university campus. Their ages ranged from 20 to 42 years. Before the individual training course started, participants were asked to fill in questionnaires on PRCS and SE. To ensure the comparability of the participants in the two conditions, a matched pairs design was employed to assign the participants randomly to either VR condition or IM condition based on their gender and PRCS scores. The training course includes three training sessions followed by a session where participants gave their presentation to a human audience. These three sessions were scheduled over three to ten working days, depending on participants' availability. Whereas the first training session and presentation in front of the human audience were scheduled on separate days, participants were allowed to schedule the second and the third session on the same day with at least one hour interval between the sessions so that people had time to reflect on and improve their presentation.

Each training session included a coaching phase and a practice phase. Figure 4.3 shows the setting for the training sessions. In the first session, the participants first completed SE, after which they started with interacting with the virtual coach, followed by rehearsing their presentation. During rehearsal, participants could present their presentation slides on a 42-inch TV (LG42lm3450). In the VR condition, before the presentation, participants chose a virtual environment, set the attitude of the virtual audience, and defined whether an interrupting event would occur during the presentation. The virtual environment was projected on a projection screen (330*250cm) by an EIKI EIP-200 projector. The participants then presented their presentations to the projected virtual audience and answered the questions asked by the virtual audience, standing with an average distance of 3.5 meters from the projection screen. In the IM condition, participants were instructed to look at the blank projection screen and to imagine an audience similar to a real situation. The participants were also instructed to think of possible reactions of the imaginary audience and possible questions from the audience. The participants were requested to keep the imaginary audience in mind when rehearsing their presentation. For either VR or IM condition, the participants needed to score their SUD every three minutes starting at the beginning of the presentation. The SUD scores were asked by the coaching system with a recorded voice automatically, and the measure procedure was the same in both conditions. Although somewhat disruptive for the presentation, this procedure is commonly used to obtain subjective anxiety levels in psychotherapy sessions [42], [43]. After finishing the whole session, the participant rated SE and PR during the presentation. The participant's answers to the questions for preparing and improving the presentation were printed out in a document and sent to the participant by email. The procedure for the second and third session was similar, except that rehearsal and coaching phase was reversed.

In the closing session, participants first rated SE and then deliver their speech in front of an audience of two people. One of the audience member was an experimenter who not only knew the condition the participants was assigned to but also had assisted the participants during training sessions, e.g. explain setup, administrated the questionnaires. The other audience member was an experimenter who neither had knowledge of participant's condition nor had interacted with participants during training sessions. Following the presentations, a question-and-answer (Q&A) phase began whereby the audience asked the participants four questions about their presentation. These questions were the same for all participants. After this, both audience members independently rated the participants' speech performance (PP) while the participants rated their SUD during the presentation, SE, and UQ. The participants' heart rates were monitored by a Bluetooth heart rate monitor (Zephyr HxM Smart) both during training sessions and closing presentation session. The training course was not part of bachelor or master program given at university. Participation was voluntary. The experiment was approved by the university ethics committee.

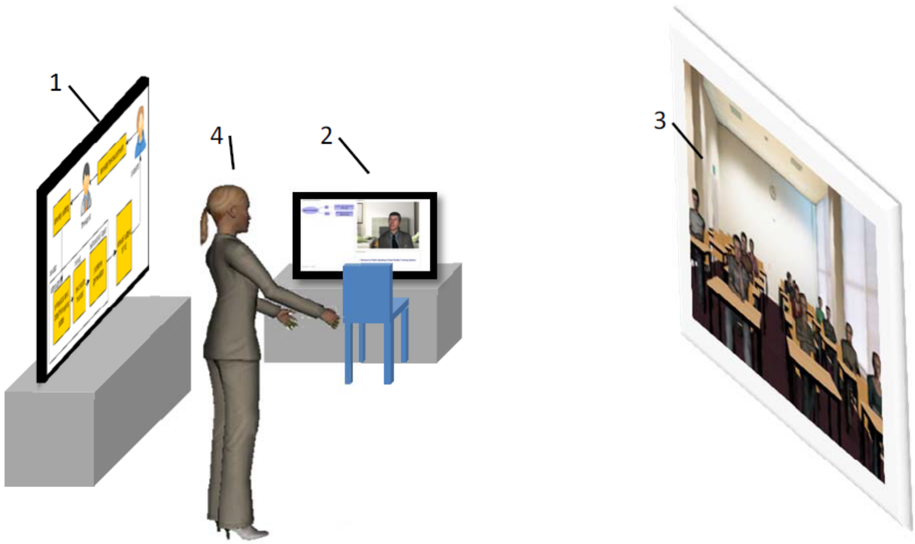


Figure 4.3: Experiment setting for the training sessions. 1 – TV for showing presentation slides; 2 – virtual coach; 3 – projection screen for virtual audiences which was left blank in imaginary condition; 4 – participant wearing a heart rate monitor.

4.4. Data Preparation and Analysis

The ratings of each rater on participants' speech performance (PP) were calculated by calculating the mean of delivery items and content items. To measure the rating consistency between the two raters, Cronbach's alpha was calculated for the delivery items (0.85), and content items (0.76). As all the ratings showed acceptable consistency, the score for delivery, and content dimensions were obtained by averaging the ratings from the two raters.

Cronbach's alpha was calculated respectively for presence response (PR), utility questionnaire on practice process, and utility questionnaire on performance outcome, ranging from 0.84 to 0.95 (Table 4.2). Due to the good consistency between their items, the mean value of the items within each questionnaire was taken as a single measure of that questionnaire. For heart rate variability (HRV), SDNN, RMSSD, and LF/HF ratio (i.e., ratio of low frequency and high frequency component) were selected as the measures following the guideline for HRV measurement [35]. The HRV measures were calculated by Kubios HRV version 2.2. The calculation of LF/HF ratio was based on Fast Fourier Transform (FFT) methods. As measures of individuals' state during the final presentation, the HRV measures during presentation and the measures during question-and-answer (Q&A) phase were respectively examined. All data were statistically analyzed in SPSS version 20 with t-tests, correlation analysis, multivariate ANOVA (MANOVA), and repeated measures ANOVA. For t-tests, bootstrap procedures were used. For repeated measure analyses, a Greenhouse-Geisser correction was used when sphericity assumption was not met.

Table 4.2: Reliability between items within the questionnaires

Questionnaire	Measuring phase / dimension	Cronbach's alpha
Presence response	Session 1	0.84
	Session 2	0.89
	Session 3	0.92
	Closing presentation	0.90
Utility questionnaire	practice process	0.91
	performance outcome	0.95

4.5. RESULTS

As eight participants (5 in VR condition, 3 in IM condition) dropped out during the experiment, the analysis only included the data of 40 participants (13 females and 27 males) who completed the whole experiment. Each condition involved 20 participants (7 females and 13 males in VR condition and 6 females and 14 males in IM condition). To check whether pre-experimental differences existed between the conditions, independent t-tests were performed. The results (Table 4.3) showed no significant differences between participants in VR condition and IM condition in PRCS, age, and self-efficacy before the first training session. Self-efficacy measured before the start of the course was found significantly correlated with self-efficacy measured at other moments (rs > 0.48, ps $\leqslant$ 0.001), and was therefore included as a covariate in the analysis of self-efficacy. However, no other measures were found to correlate significantly with either self-efficacy measured before the start of the course, PRCS or presence response (PR).

Table 4.3 shows a significantly higher presence response rating in the VR condition than in IM condition in the first training session and a similar trend for closing presentation. A mixed ANOVA on presence response with time of measurement as within-subjects factor and practicing condition (VR vs IM) as between-subjects factor found a significant main effect of measuring time ($F(3,114) = 34.25$, p <0.001) and a marginally significant main effect of condition ($F(1,38) = 3.59$, $p = 0.07$). Detailed comparisons also showed that presence response in the closing presentation was significantly higher (ps < 0.001) than presence response in other sessions.

Still, Table 4.3 also shows that no difference was found between the two training conditions with regard to the content and the way the participants presented the presentation in front of human audience.

4.5.1. ANXIETY

To study how participants' anxiety changed over the sessions, analyses were conducted on mean values of heart rates (HR) and mean values of subjective units of discomfort (SUD) during the three practice phases and the SUD score given after the closing presen-

Table 4.3: Descriptive statistics of the measures, Mean (SD), and results of independent t-tests.

Measure and phase/dimension		Condition VR	IM	Effect size Pearson's correlation, r
PRCS		12.85(5.88)	12.85(5.75)	0.00
Age		26.85(3.38)	28.10(5.56)	-0.13
Presence response	Session 1	4.68(1.19)	3.83(1.11)	0.35*
	Session 2	4.85(1.23)	4.50(1.45)	0.13
	Session 3	5.02(1.01)	4.35(1.55)	0.25
	Closing presentation	6.15(0.75)	5.63(1.08)	0.27#
HR	Session 1	87.55(9.87)	87.34(10.06)	0.01
	Session 2	89.29(10.19)	89.74(9.10)	-0.02
	Session 3	88.81(12.14)	89.84(9.90)	-0.05
	Closing presentation	101.99(20.68)	101.96(17.08)	0.00
HRV, SDNN	Closing presentation, presentation	91.61(33.68)	82.27(31.14)	0.14
	Closing presentation, Q&A	101.77(51.30)	81.76(44.60)	0.20
HRV, RMSSD	Closing presentation, presentation	112.53(51.84)	96.81(46.77)	0.16
	Closing presentation, Q&A	117.18(72.95)	92.66(64.53)	0.17
HRV, LF/HF	Closing presentation, presentation	0.99(0.47)	0.92(0.43)	0.08
	Closing presentation, Q&A	1.79(1.57)	2.04(2.53)	-0.06
SUD	Session 1	3.71(2.24)	3.84(1.78)	-0.03
	Session 2	3.85(2.28)	3.78(2.58)	0.01
	Session 3	3.55(2.51)	3.74(2.68)	-0.04
	Closing presentation	4.50(2.33)	5.15(2.52)	-0.13
Self-efficacy	When registered as a participant	6.60(1.14)	6.15(1.95)	0.14
	Before Session 1	6.10(2.55)	6.10(2.22)	0.00
	After Session 1	6.63(1.89)	6.40(1.60)	0.07
	Before Session 2	6.40(1.50)	6.30(2.20)	0.03
	After Session 2	6.85(1.35)	6.80(1.85)	0.02
	Before Session 3	7.05(1.00)	6. 08(1.82)	0.31
	After Session 3	7.40(0.94)	6.95(1.43)	0.18
	Before closing presentation	7.45(1.05)	7.00(1.78)	0.15
	After closing presentation	7.80(0.70)	7.35(1.50)	0.19
Utility questionnaire (UQ)	Process-related	5.48(0.92)	4.46(1.47)	0.38*
	Outcome-related	5.29(1.09)	3.85(1.65)	0.46**
Presentation performance (PP)	Delivery	3.09(0.42)	3.05(0.35)	0.05
	Content	3.30(0.44)	3.32(0.33)	-0.03
Length of answers (LA) in total (seconds)		173.31(56.90)	132.99(60.46)	0.32#

#$p < 0.1$, *$p < 0.05$, **$p < 0.01$.

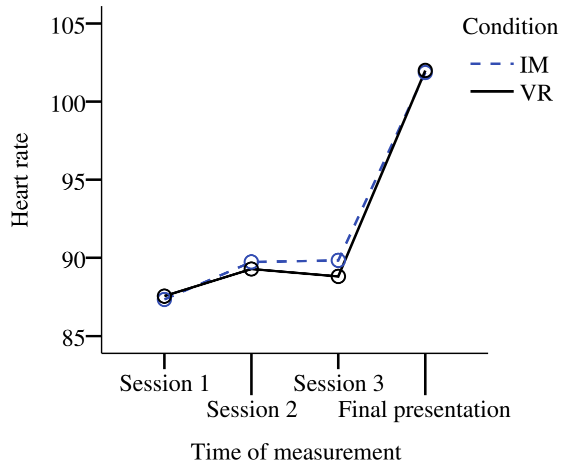
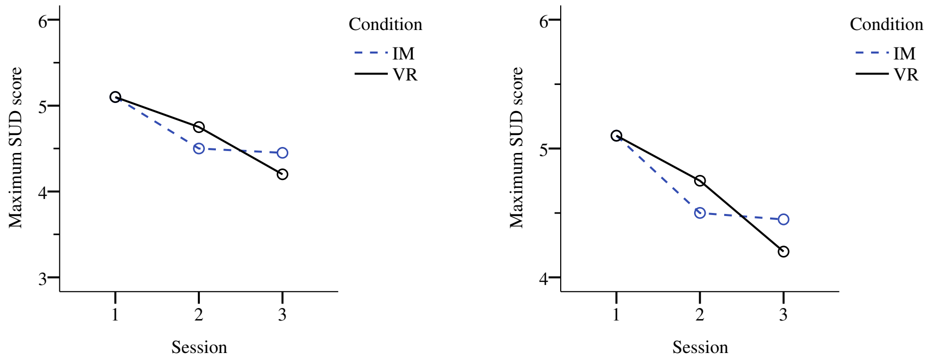


Figure 4.4: Mean of heart rates in different sessions and the final presentation.

tation. A two-way mixed ANOVA was conducted respectively on HR and SUD, with time of measurement as the within-subjects factor, and practice condition (IM versus VR) as the between-subjects factor. The results (Figure 4.4 and Figure 4.5a) showed significant main effects of time of measurement on both HR ($F(2.18, 82.83) = 18.70, p < 0.001$) and SUD ($F(1.91, 72.64) = 4.85, p = 0.01$), but no significant main effect of condition or interaction effect was found. Detailed comparisons were conducted respectively on HR and SUD between different sessions and the final presentation. The results showed that both HR (ps < 0.001) and SUD (ps < 0.02) in the closing presentation were significantly higher than those measured during the practice sessions, whereas either HR or SUD during practice sessions did not differ significantly from each other. Therefore, according to the results of both HR and SUD, the anxiety experienced in front of a real audience was much stronger than the anxiety experienced with a virtual audience or an imaginary audience. Although lower than the anxiety reported after the presentation in front of the human audience, the mean SUD scores in practice sessions were all significantly above 2.5 (one-sample t-test, ps < 0.01, suggesting that participants at least experienced some level of anxiety when practicing.

To further analyze the evoked anxiety, the maximum SUD scores reported during the practice sessions were examined. A two-way mixed ANOVA was conducted, with time of measurement as the within-subjects factor, and practice condition (IM versus VR) as the between-subjects factor. The results (Figure 4.5b) revealed only a significant main effect for time of measurement ($F(2, 76) = 5.02, p = 0.01$). Comparisons between the sessions revealed that the maximum SUD score reported in the third session was significantly lower ($p = 0.004$) than the score in the first session. Thus, the practice with either method helped individuals to lower the maximum of experienced anxiety. Additionally, no significant difference was found between the maximum SUD score reported in first training session ($t(39) = 0.65, p = 0.52$) and the SUD score reported after the final presentation in front of human audience. Both the maximum SUD of the first session and



(a) Mean scores of SUD (b) Maximum SUD score

Figure 4.5: SUD score reported in practice sessions and the final presentation.

the SUD of the final presentation were significantly higher than the mean SUD scores reported in the first three sessions ($ps < 0.001$).

To examine how the final presentation experience affected heart rate variability (HRV), a two-way mixed MANOVA was conducted on the three measures of HRV, with practice condition as the between-subjects factor and presentation-or-Q&A phase as the within-subjects factor. The results only showed a significant main effect of presentation-or-Q&A phase on the HRV measures, $F(3,21) = 3.59, p = 0.03$. Separate univariate ANOVA revealed that LF/HF ratio during the question-and-answer phase was significantly higher ($F(1,23) = 6.10, p = 0.02$) than the ratio during the presentation. This suggests that the presenters experienced more anxiety during the Q&A phase.

The lengths of the answers (LA) to the four questions from the audience were analyzed by a two-way mixed ANOVA. The training condition was included as the between-subjects factor, and the question sequence as a within-subjects factor. The results revealed a marginally significant difference ($F(1,24) = 3.07, p = 0.09$) between the conditions. As Figure 4.6 shows participants in VR condition gave longer answers than participants in the IM condition. The analysis found no significant main effect for question sequence or an interaction with training condition.

4.5.2. SATISFACTION

To examine whether participants regarded the VR practicing condition as more satisfying than the IM condition, a MANOVA was conducted on the two aspects (i.e., process and outcome) of utility questionnaire (UQ), with practice condition (IM versus VR) as the between-subjects factor. The analysis found a significant effect for practice condition, $F(2,37) = 5.17, p = 0.01$. Separate univariate ANOVA revealed that both process-related utility ($F(1,38) = 6.97, p = 0.01$) and outcome-related utility ($F(1,38) = 10.58, p = 0.002$) were rated significant higher in VR condition than in IM condition (Table 4.3).

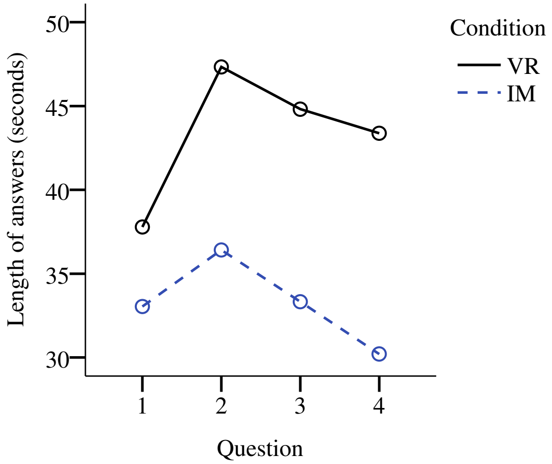


Figure 4.6: Lengths of answers to the four questions asked in the closing presentation.

4.5.3. SELF-EFFICACY

To study how self-efficacy changed over the practice sessions and the closing presentation, a mixed ANCOVA was conducted on the self-efficacy measured. The analysis included two within-subjects factors: time of measurement (i.e., three practice sessions and the final presentation), and pre-or-post measurement (i.e., whether the measure was obtained before or after the presentation). Practice condition was again included as a between-subjects factor. Finally, self-efficacy measured before the start of the course was taken as the covariate. The results (Figure 4.7) showed significant main effects for time of measurement ($F(2.19, 78.68) = 9.68, p < 0.001$), for pre-or-post measurement ($F(1,36) = 4.74, p = 0.04$), and for self-efficacy measured before the course started ($F(1,36) = 28.39, p < 0.001$), and significant interaction effects between time of measurement and self-efficacy before the course ($F(2.91, 78.68) = 6.77, p = 0.001$) and between pre-or-post measurement and self-efficacy before the course ($F(1,36) = 5.84, p = 0.02$). This suggested that delivering presentations, whenever for practice or the final presentation, help individuals to strengthen their efficacy beliefs on public speaking.

4.6. CONCLUSIONS AND DISCUSSION

This study compared two practice conditions for public speaking: practice with a virtual audience (VR) and practice with an imaginary audience (IM). The findings provided no support for the first hypothesis (H1) that VR compared to IM training conditions improved the presentation content or the way people delivered their presentation. Although because of factors such as the study's relatively small sample size, it cannot be ruled out that presentation performance would enhance after practicing with a virtual audience compared to an imaginary audience, the findings of this study suggested such an enhancement would at most be relatively small. The gain of practicing with a virtual audience however was found in the confidence of giving a presentation, specifically

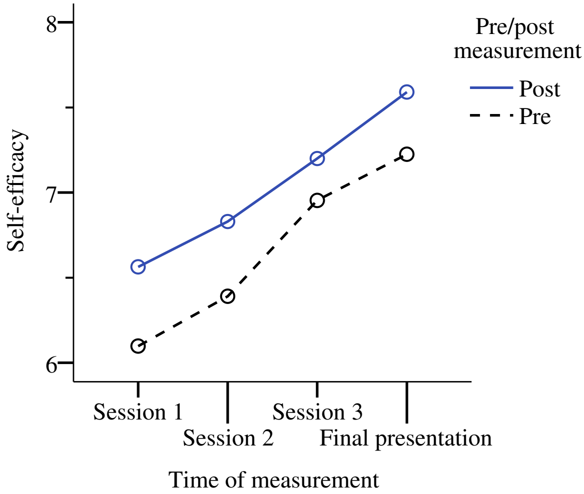


Figure 4.7: Pre-measurement versus post-measurement of self-efficacy during the course.

in answering questions. Here the analysis found a trend towards longer answers in the VR practicing condition than in the IM practicing condition. This result seems in line with other virtual reality studies that found an increase in answer length when people got positive questions and feedback from a virtual human [14], [15], or flattering comments made by virtual bystanders [16]. This finding therefore provides some support for the hypothesis that practicing with a virtual audience could reduce anxiety (H2). However, no support for this was provided by the analyses of the anxiety reported by participants or their heart rate. Interestingly, a change of the spectrum power in the heart rate variability revealed a higher level of anxiety during question-and-answer phase than the presentation phase. It suggests therefore that the question-and-answer phase is also an important part for training of public speaking skills. Still, the clearest benefit of practicing with a virtual audience was the satisfaction it gave compared with the imaginary method (H3). Participants were more positive towards training with a virtual audience, e.g. more motivating and requiring less effort, and they were also more positive about the effect this training would have on their presentation ability, e.g. less nervous and more eye contact with the audience. The presence response findings might explain this. Participants indicated to feel and behave more like presenting in front of a human audience when practicing with a virtual audience than when practicing with an imaginary audience. Although training increased participants' self-efficacy, practicing with a virtual audience or imaginary audience seemed similarly effective in accomplishing this and therefore no support for H4 was found.

Like any empirical study, this study has some limitations that should be considered. First, the study's sample size was relatively small considering the effect size observed. On the other hand, to our knowledge this study is the most extensive study nowadays comparing the impact of practicing with a virtual and an imaginary audience in the context of public speaking skills training. Second, practicing was not studied in isolation but in combination with a training provided by a virtual coach. Although in principle it was possible to study only the practicing part of training without offering additional educational support, this might be essential for helping trainees to reflect and learn from their presentation experience. The latter has been observed in several studies [44], [45] that compared guided and unguided discovery learning. Third, in this study, the reported speakers' confidence, measured with PRCS, was found only correlated with self-efficacy but not with other measures such as anxiety or performance. Thus, PRCS might not be a key measure for this study. Instead, general public speaking skills or speech performance before training might be more relevant for exploring the effectiveness of the practice methods. For example, the general speaking skills before training could be used as a criterion to assign the participants to different conditions so that unexpected differences before experiment could be controlled. Besides, it could be used as a reference to examine how much progress the participants made after the practice. Another limitation could be the lack of control over how participants made use of the virtual audience or the imaginary audience. Many participants found it very difficult to keep the imagination through the whole presentation. On the other hand, whether individuals tried to keep eye contact with the virtual audience or just read their slides was also not controlled for in this study. Besides, the cognitive workload during presentation practice could be monitored to investigate the difference between the two practice methods.

The study can be extended in many directions. First, the effectiveness of practice methods could be enhanced. For example, individuals' presentation practice could be audio or video recorded and reviewed so that they could be more aware of their own performance [46], [47]. Factors affecting the effectiveness could also be considered such as individuals' familiarity with their presentation content and motivation for success. Another direction is to enhance the virtual audience. For example, the virtual audience could behave adaptively according to the presenter's behaviour, to mimic a real-world situation. Future studies could therefore investigate the effectiveness of training with different audiences, e.g., whether an adaptive audience outperforms others in training presenters to adjust their own presentation strategy, or whether individuals control their anxiety better when practice with a bored audience, or whether a virtual audience is better than a static picture. The virtual audience could also be more intelligent. For example, using sensors for eye tracking and speech detection, the system could detect whether a person is addressing and looking at the audience. Using this information the virtual audience could adjust its behaviour, which could also be used to provide trainees with feedback on their behaviour [48], such as their gaze direction. Currently, the question and answer part at the end of the presentations included standard questions. Using speech synthesis, it would be possible that trainees could enter content-relevant questions beforehand, which a virtual audience member could pose at run time.

The main scientific contribution of the work presented is the insight that practising with a virtual audience improves training enjoyment and presentation performance in the eyes of trainees. In the end, this might be even a more instrumental asset than actually improving the presentation. When preparing a speech, people with higher levels of speech anxiety seem often reluctant to rehearse their presentation [49]. Therefore, getting them to practise would be an essential step forward as practising with audiences improves the presentation [6], reduces anxiety, and increases willingness to speak in public

[7]. In this context, a virtual audience could therefore make an important impact.

REFERENCES

[1] P. Wu, K. Yoshida, H. Suto, S. Hanita, T. Yokoi, and N. Yamaji, "The Effect of Speaker' s Behaviors on Audiences' Impressions," in *The 5th International Congress of the International Association of Societies of Design Research: IASDR 2013*, 2013.

[2] S. A. Beebe, "Eye contact: A nonverbal determinant of speaker credibility," *The Speech Teacher*, vol. 23, no. 1, pp. 21–25, Jan. 1974.

[3] J. C. McCroskey and R. S. Mehrley, "The effects of disorganization and nonfluency on attitude change and source credibility," *Speech Monographs*, vol. 36, no. 1, pp. 13–21, Mar. 1969.

[4] H. S. Park, T. R. Levine, C. Y. Kingsley Westerman, T. Orfgen, and S. Foregger, "The Effects of Argument Quality and Involvement Type on Attitude Formation and Attitude Change: A Test of Dual-Process and Social Judgment Predictions," *Human Communication Research*, vol. 33, no. 1, pp. 81–102, Jan. 2007.

[5] K. E. Menzel and L. J. Carrell, "The relationship between preparation and performance in public speaking," *Communication Education*, vol. 43, no. 1, pp. 17–26, Jan. 1994.

[6] T. Smith and A. B. Frymier, "Get 'Real': Does Practicing Speeches Before an Audience Improve Performance?," *Communication Quarterly*, vol. 54, no. 1, pp. 111–125, Feb. 2006.

[7] J. Ayres, T. Schliesman, and D. A. Sonandré, "Practice makes perfect but does it help reduce communication apprehension?," *Communication Research Reports*, vol. 15, no. 2, pp. 170–179, Mar. 1998.

[8] C. Kirschbaum, K.-M. Pirke, and D. H. Hellhammer, "The 'Trier Social Stress Test'-A tool for investigating psychobialogical stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, no. 1–2, pp. 76–81, 1993.

[9] O. Kelly, K. Matheson, A. Martinez, Z. Merali, and H. Anisman, "Psychosocial stress evoked by a virtual audience: Relation to neuroendocrine activity," *CyberPsychology & Behavior*, vol. 10, no. 5, pp. 655–62, Oct. 2007.

[10] E. Klinger, S. Bouchard, P. Légeron, S. Roy, F. Lauer, I. Chemin, and P. Nugues, "Virtual reality therapy versus cognitive behavior therapy for social phobia: A preliminary controlled study," *CyberPsychology & Behavior*, vol. 8, no. 1, pp. 76–88, 2005.

[11] M. Powers and P. Emmelkamp, "Virtual reality exposure therapy for anxiety disorders: A meta-analysis," *Journal of Anxiety Disorders*, vol. 22, no. 3, pp. 561–569, 2008.

[12] G. M. Billingsley and B. K. Scheuermann, "Using Virtual Technology to Enhance Field Experiences for Pre-Service Special Education Teachers," *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, vol. 37, no. 3, pp. 255–272, Apr. 2014.

[13] N. U. Bautista and W. J. Boone, "Exploring the Impact of TeachME^{TM} Lab Virtual Classroom Teaching Simulation on Early Childhood Education Majors' Self-Efficacy Beliefs," *Journal of Science Teacher Education*, vol. 26, no. 3, pp. 237–262, Feb. 2015.

[14] D. Hartanto, I. L. Kampmann, N. Morina, P. G. M. Emmelkamp, M. A. Neerincx, and W. P. Brinkman, "Controlling Social Stress in Virtual Reality Environments," *PLoS ONE*, vol. 9, no. 3, p. e92804, Mar. 2014.

[15] C. Qu, W. P. Brinkman, Y. Ling, P. Wiggers, and I. Heynderickx, "Conversations with a virtual human: Synthetic emotions and human responses," *Computers in Human Behavior*, vol. 34, no. May, pp. 58–68, May 2014.

[16] C. Qu, Y. Ling, I. Heynderickx, and W.-P. Brinkman, "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*, vol. 10, no. 4, p. e0125279, Jan. 2015.

[17] M. Slater, D.-P. Pertaub, C. Barker, and D. M. Clark, "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*, vol. 9, no. 5, pp. 627–33, Oct. 2006.

[18] C. Zanbaka, A. Ulinski, P. Goolkasian, and L. F. Hodges, "Social Responses to Virtual Humans: Implications for Future Interface Design," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2007, pp. 1561–1570.

[19] H. S. Wallach, M. P. Safir, and M. Bar-Zvi, "Virtual reality cognitive behavior therapy for public speaking anxiety: a randomized clinical trial.," *Behavior modification*, vol. 33, no. 3, pp. 314–38, May 2009.

[20] P. L. Anderson, M. Price, S. M. Edwards, M. a Obasaju, S. K. Schmertz, E. Zimand, and M. R. Calamaras, "Virtual reality exposure therapy for social anxiety disorder: a randomized controlled trial.," *Journal of consulting and clinical psychology*, vol. 81, no. 5, pp. 751–60, Oct. 2013.

[21] L. B. Pham and S. E. Taylor, "From Thought to Action: Effects of Process-Versus Outcome-Based Mental Simulations on Performance," *Personality and Social Psychology Bulletin*, vol. 25, no. 2, pp. 250–260, Feb. 1999.

[22] C. W. Choi, J. M. Honeycutt, and G. D. Bodie, "Effects of Imagined Interactions and Rehearsal on Speaking Performance," *Communication Education*, vol. 64, no. 1, pp. 25–44, Dec. 2014.

[23] A. Bandura, *Self-efficacy: The exercise of control.* New York: W .H. Freeman and Company, 1997.

[24] D. A. Clark and A. T. Beck, *Cognitive therapy of anxiety disorders: science and practice.* New York, New York, USA: The Guilford Press, 2010.

[25] D. Kahneman, *Attention and effort.* New Jersey: Prentice Hall, 1973.

[26] C. D. Wickens, J. Lee, Y. D. Liu, and S. Gordon-Becker, *Introduction to Human Factors Engineering*, 2nd ed. Upper Saddle River, NY, USA: Prentice-Hall Inc., 2003.

[27] N. Kang, W. P. Brinkman, M. B. van Riemsdijk, and M. A. Neerincx, "An Expressive Virtual Audience with Flexible Behavioral Styles," *IEEE Transactions on Affective Computing*, vol. 4, no. 4, pp. 326–340, Oct. 2013.

[28] N. Kang, "Posture corpus for the behavior generator [Dataset]," 2013. doi: 10.4121/uuid:d613cc9c-c10b-4c50-be50-ba8ef7885dc5.

[29] N. Kang, W.-P. Brinkman, M. B. van Riemsdijk, and M. Neerincx, "The design of virtual audiences: Noticeable and recognizable behavioral styles," *Computers in Human Behavior*, vol. 55, pp. 680–694, Feb. 2016.

[30] A. Bandura, "Guide for constructing self-efficacy scales," in *Self-efficacy beliefs of adolescents*, Information age publishing, 2006, pp. 307–337.

[31] "Oral presentation rubric." [Online]. Available: http://www.uwplatt.edu/system/files/UW-Mad Oral_presentation_rubric.pdf. Accessed: 2015-02-05.

[32] S. G. Hofmann, M. G. Newman, A. Ehlers, and W. T. Roth, "Psychophysiological differences between subgroups of social phobia.," *Journal of Abnormal Psychology*, vol. 104, no. 1, pp. 224–231, 1995.

[33] I. L. Kampmann, P. M. G. Emmelkamp, D. Hartanto, W.-P. Brinkman, B. J. H. Zijlstra, and N. Morina, "Exposure to Virtual Social Interactions in the Treatment of Social Anxiety Disorder: A Randomized Controlled Trial," *Behaviour Research and Therapy*, vol. 77, pp. 147–156, Dec. 2015.

[34] B. K. Wiederhold, D. P. Jang, S. I. Kim, and M. D. Wiederhold, "Physiological monitoring as an objective tool in virtual reality therapy," *Cyberpsychology & behavior: the impact of the Internet, multimedia and virtual reality on behavior and society*, vol. 5, no. 1, pp. 77–82, Feb. 2002.

[35] Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology, "Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, Mar. 1996.

[36] P. Nickel and F. Nachreiner, "Sensitivity and Diagnosticity of the 0.1-Hz Component of Heart Rate Variability as an Indicator of Mental Workload," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 45, no. 4, pp. 575–590, 2003.

[37] J. Wolpe, *Psychotherapy by reciprocal inhabitation*. Stanford, California: Stanford University Press, 1958.

[38] Y. Ling, H. T. Nefs, N. Morina, I. Heynderickx, and W. P. Brinkman, "A meta-analysis on the relationship between self-reported presence and anxiety in virtual reality exposure therapy for anxiety disorders.," *PloS one*, vol. 9, no. 5, p. e96144, Jan. 2014.

[39] X. Pan, M. Gillies, C. Barker, D. M. Clark, and M. Slater, "Socially anxious and confident men interact with a forward virtual woman: an experimental study.," *PloS one*, vol. 7, no. 4, p. e32931, Jan. 2012.

[40] B. G. Witmer and M. J. Singer, "Measuring Presence in Virtual Environments: A Presence Questionnaire," *Presence: Teleoperators and Virtual Environments*, vol. 7, no. 3, pp. 225–240, Jun. 1998.

[41] G. L. Paul, *Insight vs desensitization in psychotherapy*. Stanford, California: Stanford University Press, 1966.

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

[43] D. Hartanto, N. Kang, W.-P. Brinkman, I. L. Kampmann, N. Morina, P. M. G. Emmelkamp, and M. A. Neerincx, "Automatic mechanisms for measuring Subjective Unit of Discomfort," in *Annual Review of Cybertherapy and Telemedicine*, B. K. Wiederhold and G. Riva, Eds. Interactive Media Institute and IOS Press, 2012, pp. 192–196.

[44] P. A. Kirschner, J. Sweller, and R. E. Clark, "Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching," *Educational Psychologist*, vol. 41, no. 2, pp. 75–86, Jun. 2006.

[45] L. Alfieri, P. J. Brooks, N. J. Aldrich, and H. R. Tenenbaum, "Does discovery-based instruction enhance learning?," *Journal of Educational Psychology*, vol. 103, no. 1, pp. 1–18, 2011.

[46] M. Arias, P. Pando, A. Rodriguez, P. F. Miaja, A. Vazquez, M. Fernandez, and D. G. Lamar, "The Master's Thesis: An Opportunity for Fostering Presentation Skills," *IEEE Transactions on Education*, vol. 57, no. 1, pp. 61–68, Feb. 2014.

[47] T. Cochrane, "Enhancing oral presentation skills of engineering students: Technology to the rescue with the Virtual-i Presenter (ViP) [online]," in *2009 American Society for Engineering Education Annual Conference and Exposition (ASEE 2009)*, 2009.

[48] M. I. Tanveer and E. Lin, “Rhema: A Real-Time In-Situ Intelligent Interface to Help People with Public Speaking,” in *ACM international conference on intelligent user interface*, 2015.

[49] J. A. Daly, A. L. Vangelisti, and D. J. Weber, “Speech anxiety affects how people prepare speeches: A protocol analysis of the preparation processes of speakers,” *Communication Monographs*, vol. 62, no. 4, pp. 383–397, Jun. 2009.

4

5

Self-identification with a virtual experience and its moderating effect on self-efficacy and presence

Effective psychological interventions for anxiety disorders often include exposure to fearful situations. However, individuals with low self-efficacy may find such exposure too overwhelming. We created a vicarious experience in virtual reality, which enables observation of one's experience from a first person perspective without actual performance and which might increase self-efficacy. With similarities to both traditional vicarious experiences and direct experiences, the level of self-identification with the experience was hypothesized to affect self-efficacy and its relationship with direct experiences. To test this, vicarious experiences with two distinct levels of self-identification were compared in a between-subjects experiment (n = 60). After being exposed to a vicarious experience of giving lectures on elementary arithmetic in front of a virtual audience with either a high or low level of self-identification with the public speaker, participants from both conditions actively gave another lecture. The results revealed that self-identification affected people's self-efficacy after vicarious experience. They further revealed that self-identification is a moderator of (1) the correlation between perceived performance and self-efficacy, (2) the correlation between self-efficacy measured after the vicarious and the follow-up direct experience; and (3) the correlation between the sense of presence reported in the vicarious and in the follow-up direct experience. We anticipate that the first-person-perspective experiences with high-level of self-identification can be beneficial for psychotherapy or training situations where altering people's self-efficacy is desirable.

5.1. Introduction

Imagine that you are immersed in a virtual environment. You are standing in front of a speech stand and ready to give a presentation to a virtual audience. However, you do not need to speak. Instead, you hear a presentation you have never prepared. The voice sounds to you as if it is coming from you. How would you feel about such an experience? Would you feel more confident to give such a speech afterwards because you feel as if you have already successfully delivered the speech? Such experiences may help individuals with social phobia to establish their confidence and overcome their anxiety. Social phobia is a commonly occurring anxiety disorder. For example, 9.3% of the Dutch population has been estimated to suffer from social phobia during their lifetime [1], and the estimations were 12.1% for the US population [2]. As individuals with social phobia are afraid of being scrutinized and judged by others in social or performance situations, they often avoid social activities or endure extreme distress. According to Bandura's social cognitive theory [3], low self-efficacy related to coping with potential threats is the main factor that gives rise to individuals' anxiety and avoidant behaviour.

Self-efficacy is a person's subjective conviction of possessing the needed competence to cope with the demands for successfully completing a specific task. It can be enhanced by successful experiences, gained directly by individuals themselves, e.g., accomplishing tasks successfully (i.e., enactive mastery experience), or gained indirectly, i.e., by observing others' successful performance (i.e., vicarious experience) [3]. Thus, these methods are employed in psychotherapy, e.g., to help phobic individuals improve their self-efficacy in coping with potential threats, thereby eliminating their avoidant behaviour. While enactive mastery experience is considered the most influential source to establish individuals' self-efficacy, it can be problematic if individuals fail to accomplish the task, or they are even too afraid to be confronted with the situation in question in the first place. For vicarious experiences, the key to the effectiveness relies on the perceived similarity by individuals between themselves and the model in the experiences. Due to the moderating effects of the model, seeing themselves performing some sort of behaviour might be more effective than observing another conducting the behaviour in question. Video recordings can be used to observe one's own performance. However, making such recordings can be problematic because it requires the individuals to perform a certain task successfully or requires considerable video editing work for therapists to make the impression that an individual can perform an anxiety task successfully when he or she in fact is not capable to perform the task in the first place.

Virtual reality technology can provide a solution that enables an individual to experience a task from a first person perspective without actually performing it. This paper explores this approach where individuals are embodied in a virtual character that performs automatically and where they can observe this behaviour from an embodied perspective. We label this a *first-person-perspective vicarious experience*, and this experience mixes the features of a direct experience and an observed experience. On one side, the first-person-perspective vicarious experience relates to direct experiences. When a person experiences a scenario from a first person perspective, the individual may relate himself or herself to such an experience, or perceive a sense of self-identification. Hence, the individual may have the impression of performance accomplishment, which in turn may influence self-efficacy. On the other side, as an observational experience, the experience is in some aspects similar to traditional vicarious experiences. For example, the moderating effect of the model's identity may also exist on the experience's influence in self-efficacy. Therefore, this paper aims to answer what the effects of the first-person-perspective vicarious experience on self-efficacy are. Furthermore, this paper investigates the moderating effects of self-identification on the vicarious experience's relationship with both direct experiences and traditional vicarious experiences, regarding the experienced presence and the influence on individuals' self-efficacy belief.

5.2. Related Work and Hypotheses

5.2.1. Self-efficacy and its sources

Perceived self-efficacy is a person's subjective conviction of possessing the needed competence to cope with the demands for successfully completing a specific task. According to Bandura's social cognitive theory [4], self-efficacy influences people's goals and accomplishments, including how people approach challenges and goals. For example, when confronted with a challenge, people with low self-efficacy tend to avoid the situation which they believe exceeds their capability, while people with high levels of self-efficacy believe that they are capable of performing well, thereby demonstrating more effort and persistence to achieve the goal. The concept of self-efficacy is also well-recognized in other theories, such as the goal-setting theory [5] and the theory of planned behaviour [6]. In the latter case, however, it is referred to as the perceived behavioural control.

To enhance one's self-efficacy beliefs, enactive mastery experiences (i.e., performance accomplishment) are regarded as the most influential source because experiences of success or failure provide direct evidence of one's capability [7], [8]. Besides obtaining such direct experiences in real world, individuals can also obtain the experiences by actively performing specific tasks in virtual environments. For example, they can deliver a speech in front of a virtual audience [9] or answer questions as candidate in a job interview [10]. It also allows for enactive experiences with another virtual body, such as a body with another skin colour [11], a body of a different age group [12], or a body with three arms [13]. These experiences in virtual environments can affect people's self-efficacy belief as experiences in real world do. For example, people's self-efficacy can be affected by practicing a performance in virtual environments [14]. Research has further shown that effects of successful exposure to virtual environments in patients with anxiety disorders can be measured in real life [15].

Vicarious experience is another powerful source to affect self-efficacy; a person can learn by observing others or a videotaped-self performing [3]. The observational learning can be affected by the modelled performance and social comparison between the observer and the model. According to the social cognitive theory, people judge their self-efficacy partly through social comparison [3]. This judgement can be based on the performance or self-efficacy information conveyed by the modelled events. For example, when seeing a model failed repeatedly to perform a cognitive task, observers showed deteriorated self-efficacy if they were alleged to be similarly capable to the model, whereas the self-efficacy maintained high when the observers were alleged to be superior in the capability to the model [16]. When learning from vicarious experiences, besides modelled performance, people may also evaluate their own capability by comparing themselves to the model on personal characteristics such as age and gender which are assumed to be predictive of performance capabilities. For example, children have been reported to derive a stronger self-efficacy from peer modelling than observing adult models exemplifying the same task [17]. Thus, learning can be more effective, or the modelled performance is more relevant to a person, if the person perceives more similarity between oneself and the model [3]. A special case of similarity is when individuals can observe their own behaviour indirectly. For example, children learn more quickly and master more letters by watching videos of themselves instead of watching videos of someone else [18].

When individuals obtain vicarious experiences in virtual reality by observing virtual characters performing, their beliefs can also be influenced. For example, the self-efficacy became lower when observing virtual classmates praising other virtual classmates but negatively criticising the participant when answering questions in a classroom [19]. Furthermore, like vicarious experiences obtained in real life, the experiences obtained in virtual reality tend to be more influential when the virtual model is more relevant to the observers. For example, people did more physical exercises after observing a virtual lookalike, i.e. a virtual-self, jogging than observing a dissimilar virtual character jogging [20]. In a study whereby elementary children observed a virtual-self or virtual-other character swimming with whales, they developed afterwards more false memory of such swimming experience if they had observed a virtual-self [21].

Although the effects of both direct experiences and traditional experiences on self-efficacy have been well studied, it is unclear how a first-person-perspective vicarious experience affects individuals. As an experience with mixed features of direct experiences and traditional experiences, it would provide people with an observational experience, and at the same time it might also create a sense of performance accomplishment without actual performance. Hence, the principles of how direct and vicarious experiences affect self-efficacy may be generalized to the new vicarious experience. Therefore, it has the potential to be an influential source that influences self-efficacy like direct experiences and traditional vicarious experiences.

5.2.2. Self-identification in virtual reality

In virtual reality, individuals may associate themselves with certain characters, and be influenced by the experiences with the characters [22]. For example, when an illusion of body ownership is induced, light-skinned individuals showed a greater reduction in the level of racial bias after being embodied in a dark-skinned avatar than being embodied in a light-skinned avatar [11]. In the Proteus effect study [23], the appearance of avatars in online communities affected how individuals interact with others online as well as in subsequent face-to-face interactions. Furthermore, instead of being used as self-representations, the characters can also be manipulated to be identified as a different person. For example, people regarded the avatars as themselves when the avatars behaved as what they had expected whereas they identified the avatars as others when the avatars did not perform the expected behaviour [24]. In addition, as discussed earlier, the sense of self-identification has a moderating effect on the effect of traditional vicarious experiences on efficacy beliefs. Hence, as the first-person-perspective vicarious experiences are related to traditional vicarious experiences where models can be perceived or observed, the experiences are expected to affect individuals differently when different levels of associations are perceived between the models and the individuals.

5.2.3. PRESENCE

Presence is defined as “being in one place” regardless of whether the place is physical, mediated, or imagined. Witmer and Singer [25] introduced two distinct components of presence: involvement (i.e., attention side) and psychological immersion (i.e., perception of being enveloped in an environment). With an emphasis on the perception of self in a virtual environment, Biocca [26] introduced the term self-presence which represents individuals’ mental model of themselves in virtual environments when it relates to their actual body. Lee [27] also defined self-presence as a psychological state when virtual selves are experienced as the actual self in sensory and non-sensory ways. A question is, however, how presence, especially self-presence, would be experienced in a first-person-perspective vicarious.

Previous research has revealed that the first-person-perspective vicarious experiences (i.e., experiences observed in a first person perspective) are more related to one’s own experiences than traditional vicarious experiences (i.e., experiences observed in a third person perspective) when the models were identified as oneself. For example, stronger brain responses were elicited in individuals when observing an avatar committing an erroneous action in a first person perspective than observing in a third person perspective [24]. As the brain responses were also observed when individuals themselves committed errors in real life, this phenomenon suggested that the individuals regarded the errors observed in a first person perspective as committed by themselves. In a study when an avatar was slapped by another virtual character [28], participants showed greater heart rate deceleration if they observed the scene from the avatar’s perspective than from a third person perspective, and the heart rate deceleration was positively correlated with the feeling of body ownership and the feeling of being attacked or hurt. Both examples showed people perceived the avatar’s experience as their own experience, suggesting that they have experienced a certain level of self-presence. Accordingly, when a weaker sense of self-association is perceived during such experiences, the experienced presence can be expected to be also weaker. For example, when delivering speeches in a virtual environment, individuals with self-representations which were similar to themselves experienced a stronger sense of presence than those with dissimilar self-representations [9]. Therefore, the perceived self-association with the model is expected to have a moderating effect on the experienced presence in the first-person-perspective vicarious experiences.

As a result of the experienced presence, the vicarious experience may relate to direct experience regarding its effect on individuals’ efficacy beliefs. For example, individuals with specific phobias usually have a low sense of self-efficacy in coping with specific situations, and the first-person-perspective vicarious experiences have been successfully used as part of the treatment. Botella et al. [29] simulated bodily sensations such as heart palpitations and short of breath for patients suffering from panic attacks by exposing them to virtual environments in a first person perspective with the sound of heart palpitations or breathing. They compared the virtual experience treatment with another treatment, i.e., enactive mastery experience in vivo, and both treatments were found equally efficacious. Therefore, when regarded as one's own experiences, the vicarious experiences can be expected to evoke a similar effect on efficacy beliefs as direct experiences.

5.2.4. HYPOTHESES

When considering social anxiety treatment, patients with low levels of self-efficacy who find exposure to fearful social interactions too overwhelming might profit from exposure to some successful observational experience. Such an experience might increase their willingness to be exposed to direct experience and might even positively influence how they cope with the fearful situation. Against this background, four hypotheses were formulated (Figure 5.1), whereby the last two specifically focused on the relation between a vicarious experience and a follow-up direct experience. Note also that for brevity any experience discussed from now on in fact means a first-person-perspective experience. The four hypotheses are as follows:

H1: The level of self-identification affects self-efficacy after a vicarious experience.

H2: Self-identification has a moderating effect on the relationship between how people perceive the model's performance and their self-efficacy after a vicarious experience.

H3: Self-identification has a moderating effect on the correlation between presence in a vicarious and in a follow-up direct experience.

H4: Self-identification has a moderating effect on the correlation between people's self-efficacy measured after a vicarious and after a follow-up direct experience.

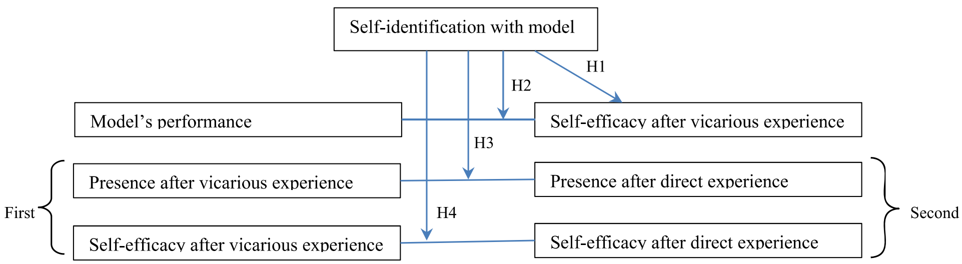


Figure 5.1: Conceptual model of the hypotheses. The solid lines represent correlational relationships between the connected constructs collected from sequential events, and the arrows stand for causal relationship. The curly brackets indicate the sequence of experiences, i.e., a vicarious experience was always followed by a direct experience.

The first hypothesis puts forwards the idea that the level of self-identification with a vicarious experience affects people's self-efficacy beliefs after such an experience. The second hypothesis proposes the underlying mechanism for this effect. The observed performance becomes an indicator for one's self-efficacy beliefs; however, the strength of such indicator is determined by how strongly people identify with the observed model. In other words, self-identification with the model determines the relevance of this performance information to form or alter self-efficacy beliefs. Furthermore, vicarious experiences share more similarities with follow-up direct experiences when people identify with the model presented in the vicarious experience. Therefore self-identification was hypothesized as a moderator on the presence correlations (H3) and on the self-efficacy correlation (H4) between these two successive events. That is, when people regard an experience more as their own experience, the sense of presence experienced in the vicarious experience corresponds more to that in a direct experience, and self-efficacy after the vicarious experience correlates more to the efficacy after a direct experience.

5.3. METHOD

To test these hypotheses, an experiment was conducted in a public speaking context. Participants were to obtain the vicarious experience by observing a job interviewee's presentation performance from the interviewee's perspective. To examine moderating effects of self-identification on the vicarious experience, a between-subjects design was employed where participants experienced one of the two distinct levels of self-identification: an experience with a low and a high level of self-identification. The two experiences were created by manipulating the virtual interviewee to be less or more like oneself. For example in the high-level condition, the virtual interviewee gave the presentations with the voice of the participant whereas the voice of the virtual interviewee sounded like another person in the low-level condition. To investigate how such a vicarious experience relates to a direct experience, a post-measurement phase was included whereby people were also asked to give a real presentation in front of a virtual audience. Figure 5.2 shows the flow of the experiment.

5

5.3.1. VIRTUAL EXPERIENCE SCENARIO AND CONDITION MANIPULATION

A public speaking scenario was created in a virtual environment. University students and staff were asked to give lectures on elementary arithmetic for school children as part of a job interview for a radio lecturer. This sample group had enough knowledge about the topic itself, yet only limited experience in giving such a lecture. In the vicarious experience, participants experienced the job interview scenario from the perspective of the interviewee. For this, they were exposed to a virtual audience [30]. The interviewee was first asked by the chair of the interview committee to introduce himself or herself and then give two lectures: one was on fraction, and the other topic was randomly assigned as multiplication or division. The order of the two lectures was chosen at random. After each lecture, the interviewee needed to answer the questions asked by the committee. An example script for the experience is shown in Appendix E. To avoid gender effect [19], [31], the virtual chair and the participant had the same gender.

A participant was assigned to be exposed to an experience with either a high or a low level of self-identification. To minimize the influence of the virtual interviewee's appearance on the participant's belief and behaviour like the Proteus effect study [23], the virtual interviewee was not fully visible. Instead, this experiment manipulated several channels described in Table 5.1 to make the virtual interviewee less or more identifiable as the participant himself or herself. In the high-level condition, the virtual interviewee spoke with the participant's own voice and used the participant's name. To make the participant notice the name, the name was mentioned nine times by the virtual committee during the vicarious experience (Appendix E), e.g., "[participant's name], would you

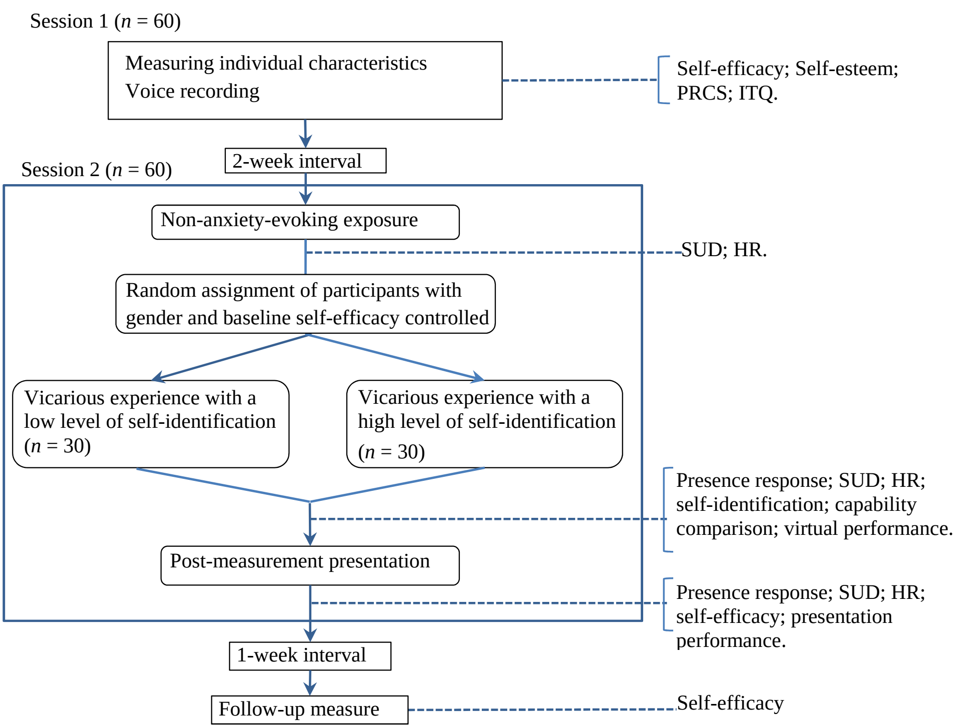


Figure 5.2: Experiment procedure and measures (listed on the right-hand side) obtained in corresponding phases. The measures are explained in Section 3.2.

please give a lecture on multiplication?" To enhance the self-identification, when looking downwards, the participant saw a gender-matched virtual body wearing a black suit, and standing with the hands holding the side of a virtual speech stand, which was the same as what he or she looked like in the real environment. In addition, the skin colour of virtual interviewee's hands matched the skin colour of the participants. In contrast, in the low-level condition, the virtual interviewee was different from the participant in all five aspects listed in Table 5.1. For example, in this condition, participants with a dark skin received a virtual body with a light skin colour. The experimental setting and screenshots of the view in the two conditions are shown in Figure 5.3.

The same scenario was employed in the post-measurement phase except that this time participants needed to deliver a real lecture instead of just observing and this time they had no virtual body. Again the topics of the lecture was on elementary arithmetic, either multiplication or division. If participants were assigned to the multiplication topic in the vicarious experience phase, they gave a lecture on division in the post-measurement phase, and the opposite was the case if they were signed to the division topic in the vicarious experience phase. While giving the lecture in this phase, participants were also requested to stand in front of the speech stand in the laboratory.

Figure 5.3: Experimental setting and screenshots of the virtual experience from the perspective of the virtual interviewee. (A) Experimental setting both in the vicarious experience and post-measurement phase. (B) A female participant's front view of the virtual experience with a female chair sitting on the left in the front row. (C) A dark-skinned male participant's top view in high-level condition. (D) A dark-skinned male participant's top view in low-level condition.

Table 5.1: The condition manipulation in the vicarious experience

Aspects	Virtual interviewee in the vicarious experience: High-level of self-identification	Virtual interviewee in the vicarious experience: Low-level of self-identification	Participant
Speaking voice	Recorded participant's own voice	Recorded voice of another participant	*
Name mentioned by the virtual committee	Participant's name	Joey for male participants and Jane for female participants	*
Arm and hand position	The same as the posture of the participant: the hands are holding the side of the speech stand in the virtual environment.	The arms and hands were positioned at the side of the virtual body.	The hands were holding the side of a speech stand.
Skin colour of hands	The same as the participant's skin.	The skin colour was much lighter or darker than the participant's skin.	*
Suit colour	Black	Light color	Dark

*No manipulation was employed.

5.3.2. MEASURES

OUTCOME VARIABLES

Self-efficacy. Following Bundura's approach [32], a one-item self-efficacy assessment was applied to measure self-efficacy. As some [33], [34] have observed, self-efficacy belief about one activity is often generalized to self-efficacy beliefs about other activities in the same domain. For example, when self-efficacy in one sport activity was strengthened, the belief in another sport activity was also enhanced, but no changes were found in the belief in for example cooking skills. Hence, self-efficacy measured in this experiment was kept within the same scenario, i.e., giving lectures on elementary arithmetic. The question was formulated as: *Please rate how certain you are that you can demonstrate to a panel of professionals that you are capable of giving radio lectures on elementary arithmetic such as subtraction and division to children aged around ten in an understandable way.* As also suggested by Bandura [32], the item was rated on a 11-point Likert scale from 0 (highly certain cannot do) to 10 (highly certain can do) with 5 (moderately can do) as the intermediate point.

Virtual performance was measured using a single scale asking participants to rate the lecture performance of the virtual interviewee (Appendix F) in the vicarious experience. Presence response. To assess how well realistic responses of people were elicited in the virtual experiences, a three-item questionnaire on presence response was adapted from the one used in Pan et al. [35] (Appendix F). It reflects presence by comparing participants' responses to what these would have been in a similar real situation – with respect to their overall behaviour, their emotional responses and their thoughts.

EXPLORATIVE AND DESCRIPTIVE MEASURES

Rosenberg self-esteem scale [36]. This is a ten-item uni-dimensional scale that measures global self-worth by measuring both positive and negative feelings about the self. All items were answered on a 4-point Likert scale ranging from 1 (strongly disagree) to 4 (strongly agree). This measure was included because the experiment involved social comparison which can influence self-esteem in a short term [37], [38].

Presentation performance. This two-item questionnaire was designed (Appendix F) for self-assessment of the presentation performance in the post-measurement phase. This measure was taken as secondary outcome to examine the effects of vicarious experience on the performance in a direct experience.

Speech length. As behavioural assessment for social anxiety [39], the lengths of the presentation and answers to the questions was taken as a measure for confidence or avoidance behaviour in the post-measurement phase. This measure was also taken as the secondary outcome. Heart rate (HR) is a physiological measure of experienced anxiety of people [40]. Participants' heart rate was monitored continuously using a Bluetooth heart rate monitor (Zephyr HxM Smart), which participants wore around the chest.

Subjective unit of discomfort (SUD) [41]. This item measures the levels of self-reported anxiety experienced by the participants. It was rated on an 11-point scale from 0 (no anxiety at all) to 10 (the highest level of anxiety that you can imagine).

Personal report of confidence as a speaker (PRCS) [42] is a 30-item self-report scale, which assesses both behavioural and affective responses to public speaking situations. The questions are answered in a true–false format, and the questionnaire score ranges from 0 (i.e., no fear of public speaking) to 30 (i.e., highest level of fear).

Immersive tendencies questionnaire (ITQ) [25]. This 18-item questionnaire measures the capability or tendency of individuals to be involved or immersed in virtual environments. Each item was rated on a seven-point semantic differential scale. ITQ rating has been associated with the level of experienced presence in a virtual environment [43].

MANIPULATION CHECK

Self-identification. To measure the identification of the experience ownership and how well people identified with the virtual interviewee, this questionnaire included six items on self-presence which was adapted from the Behm-morawitz's self-presence questionnaire [44] and two self-designed items on the perception of the virtual interviewee (Appendix F).

Capability comparison. A two-item questionnaire was designed (Appendix F) to investigate how people compared their capabilities with the observed performance of the virtual interviewee in the vicarious experience phase.

5.3.3. PARTICIPANTS

Sixty participants (24 females and 36 males) were recruited throughout the university campus. Their ages ranged from 19 to 42 years. Based on visual inspection, fifty-two participants from Europe, East Asia, Southeast Asia, West Asia, and Latin America were classified to have light skin, and eight participants from South Asia, Southeast Asia, and Latin America were classified to have dark skin.

5.3.4. PROCEDURE AND APPARATUS

The experiment included two sessions. In the first session, participants were first asked to fill in the measurements of self-efficacy, self-esteem, PRCS, and ITQ. Afterwards the speaking voices of the virtual interviewee used later in the vicarious experience were recorded. This was done for all participants. To prepare the speaking voice of the virtual interviewee, each participant was asked to read aloud a piece of text provided by the experimenter, and the reading was audio recorded with a pair of binaural microphones (Roland CS-10EM) worn by the participant. The binaural recording included spatial information of the sound sources, thereby creating a three-dimensional sound sensation for the listener. Thus, the recorded audio was supposed to sound similar to what the participant hears when he or she gives a presentation. They were instructed to read the text at a moderate speed as if they were explaining something to children in a primary school. The text included three topics on elementary arithmetic: fraction, multiplication, and division. However, the sentences from the three topics were mixed together and disarranged so that it became difficult for people to figure out the storyline, thereby minimizing the chance that the participants memorized the content. The participants were also requested to introduce themselves such as by name, occupation, and age. This audio introduction was used when the virtual interviewee introduce itself in the high-level condition.

To reduce the impact of individual difference between the two experience conditions, a matched pairs design was employed to assign the participants to either the high-level condition or low-level condition of a vicarious experience based on their gender and their self-efficacy values acquired in the first session. Hence, each condition involved 30 participants (12 females and 18 males). When assigned to the high-level condition, the recorded reading was edited to be used as the speech of the virtual interviewee. In addition to rearrangement of the audio, noticeable flaws were edited out, such as mistakes, long pauses, and hesitations, so as to make consistent and successful presentations. For the speech of the virtual interviewee in the low-level condition, a set of pre-edited readings by four other male participants and four other female participants was prepared. The noticeable flaws were also edited out. The reading records for each gender included reading by a non-native English-speaking European with a little Dutch accent, a non-native English-speaking European with a clear German accent, a native speaker with an Indian accent, and an Asian speaker with an Asian accent. The experimenter selected a gender-matched reading from the recordings of the other participants. To control for potential difference in the quality in the presentations in the vicarious experience phase and consequently the feeling of superiority, the experimenter selected recordings that matched the participant's own recording in terms of accent and pace so that the reading was perceived as comparable to the participant's own reading. Additionally, for the self-introduction in the low-level condition, the name of the virtual interviewee was edited to be Jane for female participants and Joey for males.

To ensure that the participants did not remember much about the content read, the second session was arranged to take place at least two weeks after the first session. In this session, each participant was exposed to three virtual experiences sequentially using a Sony HMZ-T2 head-mounted display (HMD) with a rotation tracker to track the participant's head orientation in three rotational degrees of freedom. The diagonal field of view of this HMD was 45 degrees. The resolution of the right and left display was 1280×720 (horizontal×vertical) pixels with a refresh rate of 60Hz. All virtual environments were displayed with stereoscopic rendering. Participants' heart rate (HR) was continuously monitored in all three exposures.

In the first exposure, a non-anxiety-evoking environment was used to obtain the baseline of the anxiety level. The participant was requested to sit and watch a short video in a virtual neutral room for five minutes [45]. SUD score during this neutral exposure was asked afterwards. Next, the participants were exposed to the vicarious experience. Before it started, they were asked to put on a black suit and stand at the speech stand in the laboratory. Their hands held the side of the speech stand in the laboratory. When ready, the vicarious experience unfolded automatically. The experience included two presentations on arithmetic of around four minutes each with a question and answer round after each presentation. The vicarious experience lasted on average 15 minutes. After the virtual experience, the participant was asked to rate the SUD, presence response, self-identification, capability comparison, virtual performance, self-efficacy, and self-esteem. Afterwards, the participants were exposed to a post-measurement phase where they needed to give a real lecture and answer the questions asked by the virtual audience in the same job interview scenario as the one in the vicarious experience. This took on average eight minutes. Afterwards, they rated the SUD, presence response, self-efficacy, and presentation performance. To check whether the vicarious experience had a long-term impact on people's self-efficacy, the participants were asked again to rate self-efficacy one week after the second session.

5.4. Data preparation and statistical analysis

Cronbach's α was calculated for the questionnaires containing multiple items, ranging from 0.77 to 0.88 (Table 5.2). Because of these acceptable levels of reliability, the mean value of the included items within each questionnaire was taken as a single measure. The data were statistically analyzed in SPSS version 20. To examine whether the two experience conditions differ in their effects on self-efficacy (H1), t-tests, repeated measures ANOVAs were performed on the outcome variables. To examine self-identification's moderating effects on the vicarious experiences (H2, H3, and H4), regression analyses with self-identification as the dichotomous moderator were firstly considered to test the hypotheses. As only linear relationships were found in the high-level self-identification condition, this violated assumptions for the dichotomous moderator variable [46], [47]. Hence, instead, the relationships hypothesized in H2, H3, and H4 were first examined by correlation analyses for each self-identification condition. The moderating effects were then investigated by comparing correlations between the two experience conditions. For t-tests, bootstrap procedures were used. For repeated measure analyses, a Greenhouse-Geisser correction was used when sphericity assumption was not met. Correlation analyses between data collected across various phases were calculated using a procedure for repeated observation data [48].

Table 5.2: Reliability between items within the questionnaires

Questionnaire and measuring phase	Removed item	Cronbach's α
Self-identification	Q6	0.83
Presence response		
Passive virtual experience	None	0.86
Post-measurement presentation	None	0.88
Capability comparison	None	0.80
Presentation performance	None	0.77

5.5. RESULTS

EACH condition involved 30 participants (12 females and 18 males). To check whether pre-experimental differences existed between the conditions, independent t-tests were performed. In addition, correlation between the measures and self-efficacy collected in the same phase were calculated. The results (Table 5.3) showed no significant differences between participants in high-level and low-level condition in self-efficacy before vicarious experience, self-esteem before vicarious experience, PRCS, and age. However, a significant difference was found in the total scale of ITQ, $t(58) = 2.51, p = 0.02$. Table 5.3 also shows that the SUD score and HR have significant negative correlation with self-efficacy rating. However, no significant correlation with self-efficacy was found in either speech length or self-esteem.

5

5.5.1. MANIPULATION CHECK

To check whether the condition manipulation of the vicarious experience was successful, a comparison was conducted between the conditions on people's self-identification. People identified the virtual interviewee significantly more ($t(58) = 5.37, p = 0.001$) as themselves in the high-level condition than in the low-level condition. Thus, people identified themselves with the high-level experience other than the low-level experience. It is important to notice that no significant difference was found between the two experience conditions in the perceived performance of the virtual interviewee (i.e., virtual performance) and how the participants compared their own capabilities with the capability of the virtual interviewee (i.e., capability comparison). This rules out an alternative explanation for affecting self-efficacy differently due to an unplanned difference in the perceived quality of the presentation by the virtual interviewee. The potential alternative explanation caused by the unplanned ITQ difference between the high-level condition and the low-level condition was also not probable because no significant correlation in either the high-level condition ($r = 0.11, n = 30, p = 0.58$) or the low-level condition ($r = 0.28, n = 30, p = 0.13$) was found between ITQ and the self-identification.

5.5.2. OVERALL ANALYSES ON SELF-EFFICACY ACROSS THE PHASES

To study how self-efficacy changed after the different experiences, a two-way mixed ANOVA was conducted on the self-efficacy measured in the recording session two weeks beforehand, after vicarious experience, after post-measurement presentation, and one

Table 5.3: Descriptive statistics of the measures, Mean (SD), and results of independent t-tests between conditions and correlation with the self-efficacy

Measure and phase		Condition: High	Condition: Low	Correlation with self-efficacy measured at the same phase [(1)], Pearson's r
PRCS		12.03(6.53)	12.47(6.39)	-0.50**
Age		26.27(4.07)	26.33(5.14)	0.16
ITQ		70.10(11.63)	62.73(11.08)*	0.13
Capability comparison		5.02(1.58)	5.02(1.75)	0.43**
Virtual performance		6.90(2.35)	7.67(1.81)	0.33*
Self-identification		6.62(1.52)	4.31(1.80)**	-0.06
SUD	Neutral room	1.03(1.19)	0.53(0.68)	
	Vicarious experience	2.47(1.48)	1.77(1.43)	
	Post-measurement presentation	5.03(2.33)	4.13(2.33)	
	Total	2.84(2.39)	2.14(2.20)	-0.45***
HR	Neutral room	75.63(14.28)	74.24(10.14)	
	Vicarious experience	84.57(14.65)	87.39(10.54)	
	Post-measurement presentation	87.85(12.80)	86.46(10.54)	
	Total	82.62(14.72)	82.56(11.94)	-0.19*
Presence response	Vicarious experience	4.58(2.10)	3.83(2.54)	
	Post-measurement presentation	6.68(2.04)	6.37(2.05)	
	Total	5.63(2.31)	5.10(2.62)	0.18
Self-efficacy	Two weeks beforehand	6.87(2.47)	7.07(2.50)	
	After vicarious experience	5.10(2.66)	6.80(2.17)*	
	After post-measurement presentation	4.43(2.66)	5.90(2.33)*	
	One week afterwards	6.37(1.61)	6.77(1.72)	
	Total	5.69(2.55)	6.63(2.22)**	
Self-esteem	Two weeks beforehand	20.87(5.49)	22.50(4.08)	
	After vicarious experience	21.03(5.89)	22.90(4.66)	
	Total	20.95(5.65)	22.70(4.35)	0.18
Presentation performance		4.38(2.41)	5.07(2.01)	0.73**
Speech length		380.61(86.66)	366.41(97.99)	-0.04

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[(1)] If a measure item was measured multiple times, the correlation was calculated within individuals between the multiple observations of the item and self-efficacy beliefs at corresponding phases.

5

week afterwards, with time of measurement as the within-subjects factor and the self-identification condition (high versus low) was included as the between-subjects factor. Significant main effects for condition ($F(1,58) = 4.30, p = 0.04$) and time of measurement ($F(2.71, 156.94) = 12.70, p < 0.001$) were found. Furthermore, a significant interaction effect was observed between condition and time of measurement ($F(2.71, 156.94) = 2.91, p = 0.04$). The mean scores are shown in Figure 5.4. Both self-efficacy beliefs after the vicarious experience and beliefs after the post-measurement presentation differed between the experience conditions (Table 5.3) and changed significantly compared with the belief measured two weeks beforehand (beforehand versus after vicarious experience, $t(59) = 2.98, p = 0.004$; after vicarious experience versus after post-measurement presentation, $t(59) = 2.29, p = 0.03$). However, no significant difference between the conditions was found either in self-efficacy measured two weeks beforehand or one week afterwards. Hence, there was no indication that the vicarious experience and post-measurement presentation had a long-term effect on self-efficacy. Furthermore, t-tests were conducted on self-efficacy in the high-level condition and low-level condition separately. Results show that self-efficacy decreased significantly ($t(29) = 3.98, p = 0.003$) in the high-level condition, while no significant change was found in the low-level condition. Thus, H1 was supported.

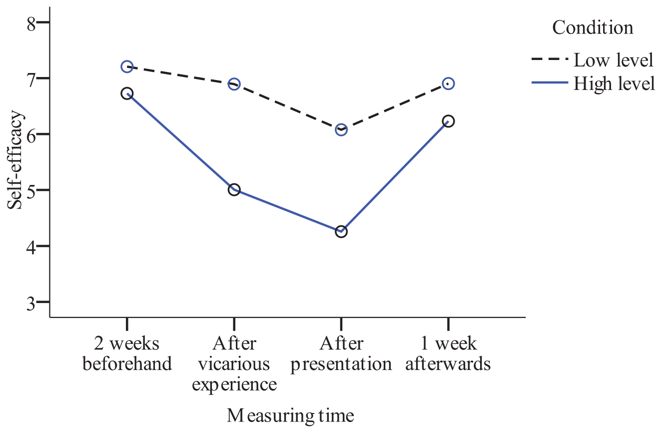
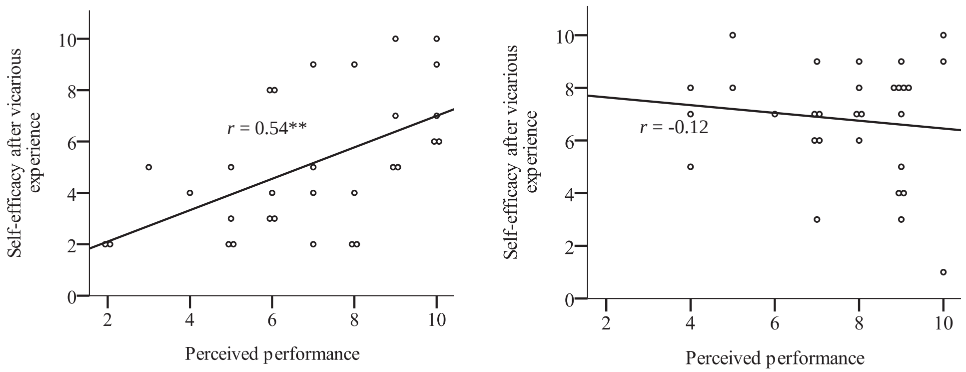


Figure 5.4: Self-efficacy measured at different times in high and low-level self-identification condition.

5.5.3. MODERATING EFFECT OF SELF-IDENTIFICATION

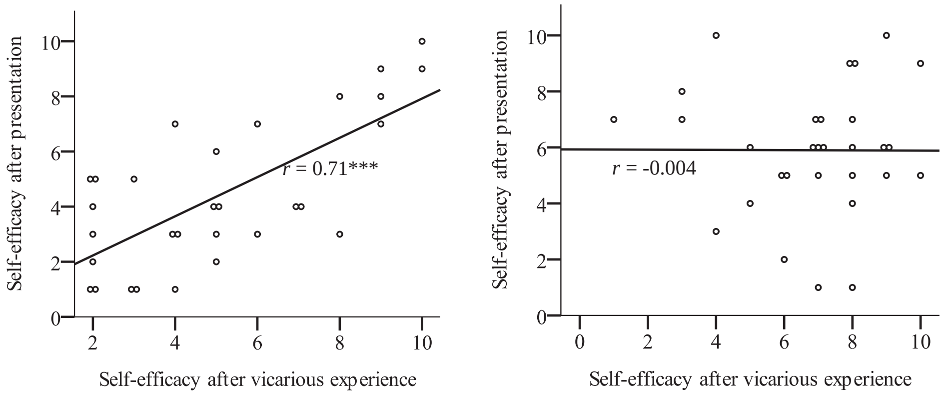
PERFORMANCE IN THE VICARIOUS EXPERIENCE AND SELF-EFFICACY AFTERWARDS

The next step was to analyze the moderating effect of self-identification on how self-efficacy was affected in the vicarious experience (H2). As shown in Table 5.3, the overall correlation was significant ($r = 0.33, n = 60, p = 0.01$) between self-efficacy after the vicarious experience and the interviewee's performance (i.e., virtual performance). Comparing the correlations between the two conditions, the correlation in the high-level condition ($r = 0.54, n = 30, p = 0.002$) was significantly higher ($z = 2.66, p = 0.01$) than the correlation in the low-level condition ($r = -0.12, n = 30, p = 0.52$). Figure 5.5 illustrates the correlations for both conditions between virtual performance and self-efficacy after the vicarious experience. Thus, participants' efficacy belief was more associated with the performance perception in the high-level condition than in the low-level condition (H2 supported).



(a) High-level condition (b) Low-level condition

Figure 5.5: Scatter plot of self-efficacy after vicarious experience and perceived performance of the virtual interviewee in the two self-identification conditions. **p < 0.01.



(a) High-level condition (b) Low-level condition

Figure 5.6: Relationship of self-efficacy before and after the post-measurement presentation in the two self-identification conditions. *** p < 0.001.

SELF-EFFICACY

The correlation between self-efficacy after vicarious experience and the one after post-measurement presentation in the high-level condition ($r = 0.71, n = 30, p < 0.001$) was significantly higher ($z = 3.28, p = 0.001$) than the correlation in the other condition ($r = -0.004, n = 30, p = 0.98$). Figure 5.6 shows the different correlations in the high-level condition and the low-level condition. The high correlation suggests that vicarious experience has a similar effect as a direct experience has on self-efficacy when the sense of self-identification is high. Thus, the level of self-identification has a moderating effect on the relationship between self-efficacy after a vicarious experience and the self-efficacy after a direct experience (H4 supported).

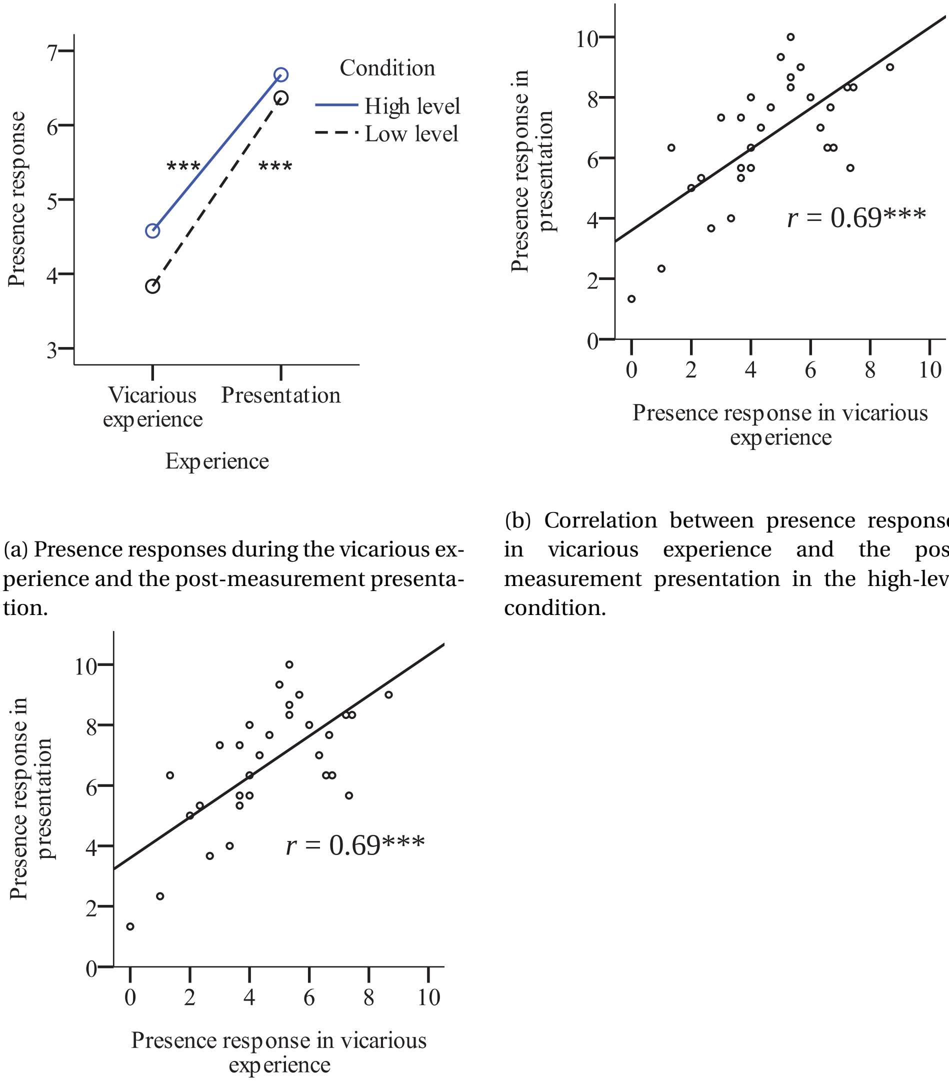
5.5.4. PRESENCE RESPONSE

A two-way mixed ANOVA was conducted on the presence responses after vicarious experience and after the post-measurement presentation, with the experience phase (vicarious experience or the post-measurement presentation) as the within-subjects factor, and self-identification condition as the between-subjects factor. Whereas no significant effect was found for self-identification condition, a significant main effect of experience phase ($F(1,58) = 49.04, p < 0.001$) was found. As Figure 5.7a reveals, the presence response in the post-measurement presentation was significantly higher than that in vicarious experience both in the high-level condition ($t(29) = -7.04, p = 0.001$) and in the low-level condition ($t(29) = -4.29, p = 0.001$). Correlations between presence response in vicarious experience and the one in the post-measurement presentation were also examined, respectively for the high-level condition and the low-level condition. As hypothesized, correlation of the high-level condition ($r = 0.69, n = 30, p < 0.001$) was significantly higher ($z = 3.04, p = 0.002$) than the correlation in the other condition ($r = 0.02, n = 30, p = 0.92$). Figure 5.7b and Figure 5.7c illustrate the different correlations in high-level condition and low-level condition. The high correlation in high-level condition indicates that the self-identification induced a similar but weaker presence experience compared with the experience induced in the post-measurement presentation. Thus, the presence reponse in the vicarious experience predicted better the presence response in the post-measurement presentation in the high-level condition than in the low-level condition. The association between ITQ and presence response was also checked respectively for the high and low-level self-identification condition. No significant correlation was observed, making it therefore unlikely that the unplanned ITQ difference between the self-identification conditions affected the reported presence response.

5

5.6. DISCUSSION AND CONCLUSIONS

BASED on the results, a number of conclusions can be drawn. First, participants in the high-level condition identified with the virtual interviewee more than participants in the low-level condition, and participants' self-efficacy changed more after the vicarious experience in the high-level condition than in the low-level condition. Thus, the results suggest that vicarious experiences with a higher level of self-identification were more able to alter self-efficacy beliefs than experiences with a lower level of self-identification, thereby supporting H1. The analyses on self-efficacy reported after the vicarious experience also revealed a significant moderator effect of self-identification on the relationship between the perceived performance and self-efficacy (H2 supported). In other words, participants who experienced a higher level of self-identification seem to have related more the performance of the virtual interviewee with their own self-



(a) Presence responses during the vicarious experience and the post-measurement presentation.

(b) Correlation between presence responses in vicarious experience and the post-measurement presentation in the high-level condition.

(c) Correlation between presence responses in the vicarious experience and the post-measurement presentation in the low-level condition.

Figure 5.7: Relationship of presence responses in vicarious experience and the post-measurement presentation. *** p < 0.001.

efficacy. When individuals experienced a weaker sense of self-identification, this performance was not regarded as exemplar for their own ability. Thus, the learning process of a traditional vicarious experience seems to be the underlying mechanism explaining the observed change in self-efficacy. Another finding was the moderating effect on the relationship between the vicarious experience and the direct experience (i.e., the post-measurement presentation) regarding the experienced presence and the effect on self-efficacy (H3 and H4). Only for participants who experienced a high level of self-identification, their beliefs established during the vicarious experience strongly correlated with their beliefs after the direct experience, and the correlation was also significant between the senses of presence in the two experience phases. However, no effect of the vicarious experience on either self-efficacy or presence was observed when the level of self-identification was low.

5

The study has a number of limitations. One limitation is the short duration and frequency of exposure to vicarious experience. Increasing its frequency might result in more lasting belief change. For example, the study by Morina et al. [49] included two exposure sessions of around one hour each and found an increase in self-efficacy at a three-month follow-up measurement. Besides, the learning effect of video self-modeling or peer-modeling can usually be observed after a few weeks of video-watching sessions [18], [50]. The limited exposure in this study might explain the observed return of the self-efficacy score after a week to pre-experimental values. Another limitation of this study is that the identity of the virtual interviewee was manipulated by controlling multiple factors (e.g., voice, skin colour and name) at the same time. Thus, it cannot be established how each factor contributed to the overall sense of self-identification. Besides these limitations, there existed an unplanned ITQ difference between the participants of the high-level condition and the low-level condition. Although previous studies [25], [51] found a significant correlation between ITQ and presence as measured by Igroup Presence Questionnaire [52], no correlation was found in this study between on one hand ITQ and on the other hand presence response, the self-identification of the virtual interviewee, and self-efficacy. Therefore, this unplanned ITQ difference can be ruled out as an alternative explanation for the observed effect. Another factor worth discussing is the manipulation of skin colour in the vicarious experience. As darker skin colour has been found to be correlated in some cases with weaker self-efficacy [53], and most of our participants were light-skinned, the participants' self-efficacy might be weakened in the low-level self-identification condition when they were embodied in a dark-skin avatar. However, a change in such a direction was not observed in this condition. Instead self-efficacy was stronger in this condition than in the high-level condition. Therefore, the manipulation of skin colour can also be ruled out as an alternative explanation for the observed effect.

Although the study succeeded in influencing people's self-efficacy belief, self-efficacy went down after the vicarious experience instead of going up which would be desirable when building someone's confidence in performing a certain task. As students and staff from a university of technology, the participants likely lacked the actual experience of giving a lecture on elementary arithmetic to ground their self-efficacy beliefs. When appraising their capability, the participants might therefore have suffered from a cognitive bias, known as the Dunning–Kruger effect [54] or overconfidence effect [55], causing the initial high self-efficacy rating [3]. After their experience in virtual reality, they might have realized that giving such lecture is more difficult than what they initially anticipated. Future work therefore might look at the effect of vicarious experiences when people unnecessarily underestimate their ability. Another issue is people's interpretations of the success of the experienced presentation. During the vicarious experience, the chair of the panel always ended the session by stating that the lecture was excellent (Appendix E). Participants might not have noticed, believed, or conformed to this judgment as this might have been too contradicting with their own belief as the social judgment theory [56] would predict. Especially when a communication message is perceived as being very different from one's own viewpoint, the message is categorized by the person as one to be rejected. The person in such case is unlikely to be persuaded, due to a contrast effect [57]. Future work might therefore explore this as another way to boost self-efficacy. Additionally, although only an association between capability comparison and self-efficacy can be argued for in this study, the relationship may be causal, as it was found in other vicarious experience studies [16]. If that is the case, future work might investigate whether self-efficacy can be improved by persuading participants to belief that they are more capable than a model that already successfully accomplished a task.

This study can be extended in many directions. First, instead of manipulating multiple factors in this study, the effectiveness of each single factor could be examined separately to achieve identification with the virtual person. As suggested by Ratan and Hasler [58], the identity of an avatar is more related to fleeting and malleable aspects such as name and appearance but less to the stable characteristics such as race and gender. Therefore, identifiable channels such as voice might be of interest in manipulating the avatar or model identity. Second, future research could examine the extent to which the vicarious experience can be used effectively to assist psychological interventions for individuals with mental disorders. For example, individuals with anxiety disorders might profit from this approach by applying the vicarious experience as a first step to increase motivation to participate in exposure sessions of direct experience in virtual reality and further in real life. Third, the use of vicarious experience in skills training could also be of value, such as in helping people to visualize their future performance and master a task in a short time.

In conclusion, our method succeeded in influencing people's self-efficacy belief by a vicarious experience obtained in virtual reality. Another main contribution of the study is the insight of the underlying mechanism that might govern people's self-efficacy. When the virtual model in the vicarious experience is more strongly identified with the observers themselves, the performance of the virtual model becomes a better predictor of the observers' self-efficacy. The mechanism seems closely related to how traditional vicarious experiences affect self-efficacy. Future research needs to investigate the extent to which this new virtual vicarious experience has the potential to benefit psychological interventions where the belief in one's own ability is essential.

REFERENCES

[1] R. de Graaf, M. Ten Have, C. van Gool, and S. van Dorsselaer, "Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental

Health Survey and Incidence Study-2," *Soc Psychiatry Psychiatr Epidemiol*, vol. 47, no. 2, pp. 203–213, 2012.

[2] R. C. Kessler, P. Berglund, O. Demler, R. Jin, K. R. Merikangas, and E. E. Walters, "Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication," *Archieves of General Psychiatry (Arch. Gen. Psychiatry)*, vol. 62, no. June, pp. 593–602, 2005.

[3] A. Bandura, *Self-efficacy: The exercise of control.* New York: W .H. Freeman and Company, 1997.

[4] A. Bandura, "Cultivate self-efficacy for personal and organizational effectivenes-sand," in *Handbook of principles of organization behavior*, 2nd ed., E. A. Locke, Ed. New York: Wiley, 2009, pp. 179–200.

[5] E. a. Locke and G. P. Latham, "Building a practically useful theory of goal setting and task motivation: A 35-year odyssey.," *American Psychologist*, vol. 57, no. 9, pp. 705–717, 2002.

[6] I. Ajzen, "The theory of planned behavior," *Organizational Behavior and Human Decision Processes*, vol. 50, no. 2, pp. 179–211, Dec. 1991.

[7] A. Garcia-Palacios, C. Botella, H. Hoffman, and S. Fabregat, "Comparing acceptance and refusal rates of virtual reality exposure vs. in vivo exposure by patients with specific phobias," *CyberPsychology & Behavior*, vol. 10, no. 5, pp. 722–724, Oct. 2007.

[8] G. Robillard, S. Bouchard, S. Dumoulin, T. Guitard, and E. Klinger, "Using virtual humans to alleviate social anxiety: preliminary report from a comparative outcome study.," *Studies in health technology and informatics*, vol. 154, pp. 57–60, Jan. 2010.

[9] L. Aymerich-Franch, R. F. Kizilcec, and J. N. Bailenson, "The relationship between virtual self similarity and social anxiety," *Frontiers in Human Neuroscience*, vol. 8, p. 944, Nov. 2014.

[10] D. Hartanto, I. L. Kampmann, N. Morina, P. G. M. Emmelkamp, M. A. Neerincx, and W. P. Brinkman, "Controlling social stress in virtual reality environments," *PLoS ONE*, vol. 9, no. 3, p. e92804, Mar. 2014.

[11] T. C. Peck, S. Seinfeld, S. M. Aglioti, and M. Slater, "Putting yourself in the skin of a black avatar reduces implicit racial bias," *Consciousness and Cognition*, vol. 22, no. 3, pp. 779–787, 2013.

[12] D. Banakou, R. Groten, and M. Slater, "Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 31, pp. 12846–51, Jul. 2013.

[13] A. S. Won, J. Bailenson, J. Lee, and J. Lanier, “Homuncular flexibility in virtual reality,” *Journal of Computer-Mediated Communication*, vol. 20, no. 3, pp. 241–259, May 2015.

[14] G. Robillard, S. Bouchard, S. Dumoulin, T. Guitard, and E. Klinger, “Using virtual humans to alleviate social anxiety: Preliminary report from a comparative outcome study,” *Studies in Health Technology and Informatics*, vol. 154, pp. 57–60, 2010.

[15] N. Morina, H. Ijntema, K. Meyerbröker, and P. M. G. Emmelkamp, “Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments,” *Behaviour Research and Therapy*, vol. 74, pp. 18–24, Nov. 2015.

[16] I. Brown and D. K. Inouye, “Learned helplessness through modeling: the role of perceived similarity in competence,” *Journal of Personality and Social Psychology*, vol. 36, no. 8, pp. 900–908, 1978.

[17] D. H. Schunk and A. R. Hanson, “Peer models: Influence on children’s self-efficacy and achievement.,” *Journal of Educational Psychology*, vol. 77, no. 3, pp. 313–322, 1985.

[18] A. Marcus and D. A. Wilder, “A comparison of peer video modeling and self video modeling to teach textual responses in children with autism,” *Journal of Applied Behavior Analysis*, vol. 42, no. 2, pp. 335–341, Jan. 2009.

[19] C. Qu, Y. Ling, I. Heynderickx, and W.-P. Brinkman, “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*, vol. 10, no. 4, p. e0125279, Jan. 2015.

[20] J. Fox and J. N. Bailenson, “Virtual self-modeling: The effects of vicarious reinforcement and identification on exercise behaviors,” *Media Psychology*, vol. 12, no. 1, pp. 1–25, Mar. 2009.

[21] K. Y. Segovia and J. N. Bailenson, “Virtually true: Children’s acquisition of false memories in virtual reality,” *Media Psychology*, vol. 12, no. 4, pp. 371–393, Nov. 2009.

[22] M. Slater and M. V. Sanchez-Vives, “Transcending the self in immersive virtual reality,” *Computer*, vol. 47, no. 7, pp. 24–30, Jul. 2014.

[23] N. Yee, J. N. Bailenson, and N. Ducheneaut, “The Proteus effect: implications of transformed digital self-representation on online and offline behavior,” *Communication Research*, vol. 36, no. 2, pp. 285–312, Jan. 2009.

[24] E. F. Pavone, G. Tieri, G. Rizza, E. Tidoni, L. Grisoni, and S. M. Aglioti, “Embodying others in immersive virtual reality: Electro-cortical signatures of monitoring the errors in the actions of an avatar seen from a first-person perspective.,” *The Journal*

of neuroscience: the official journal of the Society for Neuroscience, vol. 36, no. 2, pp. 268–79, Jan. 2016.

[25] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence: Teleoperators and Virtual Environments*, vol. 7, no. 3, pp. 225–240, Jun. 1998.

[26] F. Biocca, "The Cyborg's dilemma: Progressive embodiment in virtual environments," *Journal of Computer-Mediated Communication*, vol. 3, no. 2, pp. 1–25, Jun. 1997.

[27] K. M. Lee, "Presence, Explicated," *Communication Theory*, vol. 14, no. 1, pp. 27–50, Feb. 2004.

[28] M. Slater, B. Spanlang, M. V Sanchez-Vives, and O. Blanke, "First person experience of body transfer in virtual reality.," *PloS one*, vol. 5, no. 5, p. e10564, Jan. 2010.

[29] C. Botella, A. García-Palacios, H. Villa, R. M. Baños, S. Quero, M. Alcañiz, and G. Riva, "Virtual reality exposure in the treatment of panic disorder and agoraphobia: A controlled study," *Clinical Psychology & Psychotherapy*, vol. 14, no. 3, pp. 164–175, May 2007.

[30] N. Kang, W. P. Brinkman, M. B. van Riemsdijk, and M. A. Neerincx, "An expressive virtual audience with flexible behavioral styles," *IEEE Transactions on Affective Computing*, vol. 4, no. 4, pp. 326–340, Oct. 2013.

[31] L. Carli, "Gender differences in interaction style and influence.," *Journal of Personality and Social Psychology*, vol. 56, no. 4, pp. 565–576, 1989.

[32] A. Bandura, "Guide for constructing self-efficacy scales," in *Self-efficacy beliefs of adolescents*, Information age publishing, 2006, pp. 307–337.

[33] A. Bandura and N. E. Adams, "Analysis of self-efficacy theory of behavioral change," *Cognitive Therapy and Research*, vol. 1, no. 4, pp. 287–310, Dec. 1977.

[34] I. Scopelliti, S. Botti, and C. Donato, "Vicarious control: Exposure to mastery and perceived self-efficacy," in *the 40th International Research Conference in Marketing*, 2013.

[35] X. Pan, M. Gillies, C. Barker, D. M. Clark, and M. Slater, "Socially anxious and confident men interact with a forward virtual woman: an experimental study.," *PloS one*, vol. 7, no. 4, p. e32931, Jan. 2012.

[36] M. Rosenberg, *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press, 1965.

[37] M. Argyle, *Social encounters: Contributions to social interaction*. Aldine Transaction, 2008.

[38] M. Bari, "Measuring short-term media effects upon women's self esteem," 2007.

[39] S. G. Hofmann, M. G. Newman, A. Ehlers, and W. T. Roth, “Psychophysiological differences between subgroups of social phobia.,” *Journal of Abnormal Psychology*, vol. 104, no. 1, pp. 224–231, 1995.

[40] B. K. Wiederhold, D. P. Jang, S. I. Kim, and M. D. Wiederhold, “Physiological monitoring as an objective tool in virtual reality therapy,” *Cyberpsychology & behavior: the impact of the Internet, multimedia and virtual reality on behavior and society*, vol. 5, no. 1, pp. 77–82, Feb. 2002.

[41] J. Wolpe, *Psychotherapy by reciprocal inhabitation*. Stanford, California: Stanford University Press, 1958.

[42] G. L. Paul, *Insight vs desensitization in psychotherapy*. Stanford, California: Stanford University Press, 1966.

[43] Y. Ling, H. T. Nefs, W.-P. Brinkman, C. Qu, and I. Heynderickx, “The relationship between individual characteristics and experienced presence,” *Computers in Human Behavior*, vol. 29, no. 4, pp. 1519–1530, Jul. 2013.

[44] E. Behm-morawitz, “Mirrored selves: The influence of self-presence in a virtual world on health , appearance , and well-being,” *Computers in Human Behavior*, vol. 29, no. 1, pp. 119–128, 2013.

[45] B. Busscher, D. De Vliegher, Y. Ling, and W.-P. Brinkman, “Physiological measures and self-report to evaluate neutral virtual reality worlds,” *Journal of CyberTherapy & Rehabilitation*, vol. 4, no. 1, pp. 15–25, 2011.

[46] H. Aguinis, *Regression analysis for categorical moderators*. New York, NY: Guilford Press, 2004.

[47] A. Field, *Discovering statistics using SPSS*, 3rd ed. SAGE Publications, 2009.

[48] J. M. Bland and D. G. Altman, “Statistics notes: Calculating correlation coefficients with repeated observations: Part 1—correlation within subjects,” *BMJ*, vol. 310, p. 446, 1995.

[49] N. Morina, W.-P. Brinkman, D. Hartanto, I. L. Kampmann, and P. M. G. Emmelkamp, “Social interactions in virtual reality exposure therapy: A proof-of-concept pilot study.,” *Technology and health care: official journal of the European Society for Engineering and Medicine*, vol. 23, no. 5, pp. 581–9, Sep. 2015.

[50] R. Montgomerie, “Video self-modelling as an intervention for oral reading fluency in year 3 students,” Massey University, 2011.

[51] Y. Ling, W.-P. Brinkman, H. T. Nefs, C. Qu, and I. Heynderickx, “Effects of stereoscopic viewing on presence, anxiety, and cybersickness in a virtual reality environment for public speaking,” *Presence: Teleoperators and Virtual Environments*, vol. 21, no. 3, pp. 254–267, 2012.

[52] T. Schubert, F. Friedmann, and H. Regenbrecht, “The experience of presence: factor analytic insights,” *Presence Teleoperators and Virtual Environments*, vol. 10, no. 3, pp. 266–281, 2001.

[53] M. S. Thompson and V. M. Keith, “THE BLACKER THE BERRY: Gender, skin tone, self-esteem, and self-efficacy,” *Gender & Society*, vol. 15, no. 3, pp. 336–357, Jun. 2001.

[54] J. Kruger and D. Dunning, “Unskilled and unaware of it: How difficulties in recognizing one’s own incompetence lead to inflated self-assessments.,” *Journal of Personality and Social Psychology*, vol. 77, no. 6, pp. 1121–1134, 1999.

[55] D. A. Moore and P. J. Healy, “The trouble with overconfidence.,” *Psychological review*, vol. 115, no. 2, pp. 502–17, Apr. 2008.

[56] C. I. Hovland and M. Sherif, *Social judgment: Assimilation and contrast effects in communication and attitude change*. Westport: Greenwood, 1965.

[57] E. Griffin, *A first look at communication theory*, 8th ed. New York: New York: McGraw Hill, 2011, pp. 194–204.

[58] R. Ratan and B. Hasler, “Exploring self-presence in collaborative virtual teams.,” *PsychNology Journal*, vol. 8, no. 1, pp. 11–32, 2010.

6

Discussion and Conclusions

The research presented in this thesis implemented an expressive virtual audience, validated individuals' perception of the audience's expressiveness, investigated the application of virtual audiences in the context of public speaking training, and tested the effects of indirect virtual experiences of public speaking on individuals' self-efficacy belief about public speaking. The studies were designed to answer the main research question:

> *Within public speaking scenarios, how can an expressive virtual audience be created and how do the experiences with a virtual audience affect the presenter?*

From the main research question, four hypotheses were formulated (the first and second focus on the creation and perception of virtual audience's expressive behaviors, and the third and fourth focus on their effects on the presenter).

H1. A parameterized audience model can generate virtual audiences with expressive behaviour.

H2. People can recognize People can recognize different styles of the expressive behaviours generated by the audience model.

H3. People are more satisfied about practicing with a virtual audience than practicing with an imaginary audience.

H4. The level of self-identification affects self-efficacy after a first-person-perspective vicarious experience.

The first hypothesis was supported by the development of a parameterized agent model that was proved to be able to generate flexible expressive behaviors for the virtual audience. To test the second hypothesis, a series of perception experiments were conducted and showed that people could distinguish the model-based audience styles. The third and the fourth hypotheses focus on the effects of the model-based audience on presenters. People found practicing with a virtual audience more satisfying than practicing with an imaginary audience in our experiments. Furthermore, findings also showed that both direct and indirect experiences as a public speaker in virtual reality influenced self-efficacy beliefs in public speaking, and that direct experiences also affect people's perceptions of speech performance. The conclusions in this thesis are structured by an examination of the arguments for these four hypotheses.

H1: the generation of flexible expressive audience behavior

Support for the first hypothesis was established by implementing a virtual audience using a parameterized agent model and conducting perception studies of the virtual audience's behaviour, which investigated whether individuals can recognize the generated expressive behaviours. A parameterized audience model was created based on statistical models abstracted from observations of real audiences. The audience's behavior was controlled by model parameters that defined virtual humans' moods, attitudes, and personalities. Employing these parameters as predictors, the audience model significantly predicted variations in the audience behavior. To investigate if people could recognize the designed behavioral styles generated by this model, 12 audience styles were evaluated by two groups of participants. One group (n = 22) was asked to describe the virtual audience freely, and the other group (n = 22) was asked to rate the audiences on eight audience parameter dimensions. The results indicated that people could recognize different audience attitudes and that they even could perceive the different degrees of certain audience attitudes.

H2: recognizable behavioral audience styles
Support for the second hypothesis was established by conducting four studies on a simulated audience and its behavioral models. (I) To identify perceivable audience characteristics, a paired comparison perception experiment on the simulated audience was conducted with 24 participants. They were asked to compare pairs of virtual audiences with different characteristic settings to identify which audience characteristic resulted in noticeable audience behavior differences. The results showed that people can perceive changes in some of the mood, personality, and attitude parameters. (II) To investigate people's understanding of audience styles, a design experiment was conducted in which people (n = 24) were invited to use the perceivable characteristics to design audience behavior for a set of public speaking scenarios. By clustering the audience scenarios based on the similarity of the characteristic settings, five generic audience styles were obtained. (III) The validation of the five generic audience styles was performed in a perception study. Videos of virtual audiences were made for each style, and 100 individuals were invited to match audience style description to each video. The results suggested that a dominating characteristic of an audience people perceive is whether or not the audience is attentive or inattentive, and people can distinguish between various behavioral audience styles such as interested-enthusiastic style, critical-concerned style, and anxious-threatened style. (IV) To find out the key behavior of the audience styles, the behavioral model of the simulated audience was examined and several specific audience behaviors were identified as typical for these styles. Thus, people do have some consistent expectations about the behavior of an audience, and they do recognize the behavioral audience styles.

H3: speakers' satisfaction with practicing with a virtual audience
The third hypothesis was tested by organizing a public speaking course in an experimental setting. The course consisted of three training sessions and one closing presentation in front of a human audience. In each training session, participants obtained direct experiences as a public speaker by practicing their own presentations in front of a virtual audience projected on a screen. This practice method was compared with another method: practice with an imaginary audience whereby participants were instructed to practice their presentations while imagining the presence of an audience. In the closing presentation, participants delivered their presentations to a human audience. The study employed a between-subjects design, for which 20 participants were recruited for each practicing condition. Participants' performance was assessed in the closing session. The results showed that individuals seemed to benefit more from a virtual audience than an imaginary audience in reducing public speaking anxiety, specifically in a question-and-answer session. The clearest benefit of practicing with a virtual audience was the satisfaction it gave. Participants were more positive towards training with a virtual audience regarding both the training process and its effect on their presentation ability.

H4: the moderating effect of self-identification in vicarious experience from a first person perspective on self-efficacy beliefs

The fourth hypothesis was examined in an empirical study whereby an indirect experience of public speaking was created in VR, which enables observation of a person giving lectures on elementary arithmetic in front of a virtual audience from the speaker's perspective. To explore the moderating effect of self-identification on such vicarious experiences, the high-self-identification experiences were compared with the experiences with a low level of self-identification in a between-subjects experiment (n = 60). In the experiment, participants were first exposed to an experience with either a high or a low level of self-identification. Afterwards, participants from both conditions actively gave another lecture on elementary arithmetic. The results revealed that vicarious experience with a higher level of self-identification was more effective in modifying people's self-efficacy than an experience with a lower level. Additionally, like traditional vicarious experiences, the sense of self-identification also had a moderating effect on how self-efficacy is affected. That is, only when individuals related themselves to the experience, the perceived performance of the virtual speaker correlated significantly with people's self-efficacy measured after the vicarious experience. Next, the moderating effect also existed on the association between the vicarious experience and a direct experience. Only when a high level of self-identification was present, the vicarious experience was similar to a direct experience regarding to the experienced presence and their effects on self-efficacy scores. However, any changes in self-efficacy during the experiment were no longer observed at a two-week follow-up measurement. Therefore, the first person perspective vicarious experiences as a public speaker are effective in modifying individuals' self-efficacy beliefs in public speaking at least for a short term, and the underlying mechanism that might govern people's self-efficacy beliefs seems closely related to how vicarious experiences affect self-efficacy.

6.1. Limitations

To appreciate the work presented in this thesis, it is also important to consider its limitations. Some of the limitations are caused by the research approach applied. First, all empirical studies, with the exception of the online perception validation study (Chapter 4), recruited mainly university students. This therefore might limit generalisation of the findings to other population groups. Another limitation is that this thesis mainly employed a fixed research approach, i.e. quantitative approach, whereby various phenomena and outcomes were studied to test hypotheses. However, a flexible research approach, i.e. qualitative approach, would also address questions on why people perceive an audience in a certain way, and why presenters want to practise in front of a virtual audience.

Logistical and time constrains also caused limitations. For example, the design of the virtual audience was based on observations of a limited group of students in a controlled classroom setting, thereby obtaining limited audience behaviour with a social-economic background of western university students. Also observations were done in equal power-relationship between audience and presenter. Therefore, generalization of the audience behavior towards an unequal power-relationship might be limited. The audience model is also limited when considering interaction between presenters and their audience. For example, the audience could provide feedback on speakers' performance [1]. Besides, the effects of the relationship between the audience members could also be considered in the behavioural models of the virtual audience.

Several key factors were not considered when studying the direct experiences as a public speaker in virtual reality. First, individuals' motivation was not investigated. However, in a learning context, individuals are more likely to improve their competence when they strive for positive outcomes [2]. In contrast, if they are not motivated, the effect of training can be weakened. As Schunk [2] also pointed out, individuals could be motivated when they are aware of their lack of requisite knowledge and skill competence.

The study on the first-person-perspective vicarious experiences also has some limitations. First, the experiment only included a relatively short vicarious experience, failing therefore to establish any long-term effect. Making generalisation about longer and more frequent vicarious experiences is therefore not possible. Still long-term effects have been found for direct virtual experiences. For example, Morina et al. [3] found an increase in self-efficacy at a three-month follow-up measurement after two exposure sessions of around one hour each. Thus, increasing the exposure frequency might result in more lasting belief change. Second, participants were not familiar with the presentation topic in the vicarious experience, i.e., elementary arithmetic. Hence, they might initially have overestimated their capabilities thereby nullifying the potential enhancing effect of vicarious experience on their self-efficacy beliefs.

6.2. CONTRIBUTIONS

THIS thesis studied the implementation of an expressive virtual audience, and explored the effects of public speaking experiences in virtual reality on presenters. The scientific and practical contributions are discussed below.

6.2.1. SCIENTIFIC CONTRIBUTIONS

IMPLEMENTATION OF AN AUDIENCE MODEL

This thesis proposes an audience model that generates flexible expressive behavioral styles by adjusting agent parameters. Among the virtual audiences developed for various applications, the audiences' behavior are usually designed based on psychological knowledge and literature [4], [5], whereas this audience model is, to my knowledge, the first that generates flexible expressive behaviors based on statistical models abstracted from observations. In addition, as most studies on the generation of expressive behavior of virtual characters [6–8] focus on the behavior of virtual individuals, the proposed behavioral model is the first one designed for audiences in public speaking scenarios. This is an important step towards providing users with a flexible and dynamic virtual environment in which they can be exposed to a virtual audience, for example, as part of a psychological stress test procedure, training, or psychotherapy. In those settings, virtual audiences often function as key stimulus material. Thus, validation studies presented in the thesis are vital as they provide the foundation for drawing any valid conclusions later on about people's behavior, emotions, and attitudes when they are exposed to these virtual audiences. Besides, the audience model was built using a statistical approach based on observations of real audiences in public speaking situations. It provides a method for simulating expressive behavior of audiences and a coding scheme for posture observation. Moreover, a corpus of audience behavior [9] showing different attitudes in public speaking situations is provided so that it can be used in future studies on audience behavior.

Public speaking experiences with a virtual audience

To our knowledge the public speaking training study presented in this thesis is, to the best of my knowledge, the first study comparing the impact of practicing with a virtual and an imaginary audience in the context of public speaking skills training. The main scientific contribution of the work presented is the insight that practicing with a virtual audience improves training enjoyment and presentation performance in the eyes of trainees. Besides, a virtual audience is found similarly effective in evoking social anxiety to an imaginary audience, and this finding is consistent with the findings by Kelly et al [10].

The first-person-perspective vicarious experiences succeeded in influencing people's self-efficacy belief. Manipulating mainly the voice of a virtual person influenced the identification of this virtual person. This provides an effective and easy way to manipulate the sense of self-identification. Another main contribution of the study is the insight of the underlying mechanism that might govern people's self-efficacy beliefs. As the sense of self-identification increases, the observed performance becomes a predictor of self-efficacy beliefs. The mechanism seems closely related to how traditional vicarious experiences affect self-efficacy [11]. It shows the potential of this first-person-perspective vicarious experience to benefit psychological interventions where the belief in one's own ability is essential.

6.2.2. Practical Contributions

Public speaking skills are important because we have to speak in public from time to time, e.g., talking about our research at a conference, making a speech at a friend's wedding, or making a proposal in a business meeting. This thesis gives insight as to how a virtual audience can be developed and how the virtual audience can be used for the training of public speaking skills.

Development of expressive virtual characters

A parameterized audience model is built in this thesis that generates flexible expressive behavioral styles by adjusting agent parameters of mood, attitude, and personality. The model employs a common framework for multi-agent systems, and therefore can be generalized to the creation of expressive virtual humans for a flexible and dynamic virtual environment. The validation studies show the potential of the virtual audience's expressiveness. That is, characteristics of audience members such as their mood, personality, and attitude are perceivable in the body language of a virtual audience. Thus, developers can use this insight to establish virtual audiences that can exhibit various behavioral styles.

The finding of at least five underlying audience styles among the 21 public scenarios suggests that the five styles could represent a large variety of audiences, which could occur in various public speaking scenarios. Thus, by implementing only five audience styles, designers would be able to construct many more different social settings with an audience, and users would have opportunities to experience more variations of social settings. Besides, as an audience's attentiveness was found as a dominating perceivable characteristic, it is an important characteristic to be mentioned and considered when describing or designing an audience. Additionally, the specific postures and behavioral patterns found in the five audience styles may help designers to develop virtual audiences with noticeable and recognizable behavioral styles. The findings can also be generalized to the design of individual virtual characters acting as listeners. Specifically, to design expressive virtual listeners, their behavior should vary on the following aspects: head and gaze direction, facial expression, frequency of bodily movements, reaction to disturbing events, and postural features such as openness, relaxation, and fidgets. These findings give designers directions on how to modulate these behaviors to create a single listening individual as well as a complete virtual audience.

ANXIETY-EVOKING APPLICATIONS

The results show both direct and indirect experiences of public speaking with a virtual audience evoke anxiety in individuals. Compared with direct experiences, i.e., speaking directly in front of a virtual audience, the first-person-perspective vicarious experiences seem to have similar but weaker effects on individuals' anxiety. Specifically, individuals reported higher anxiety for direct experiences but their heart rates during direct experiences did not differ from those during the vicarious experiences. The findings that there are more ways to evoke individuals' social anxiety may benefit applications that need to elicit social anxiety in individuals. For example, to study the effect of social stress on individuals, moderate social stress can be evoked during the vicarious experiences whereas more intensive stress can be evoked by direct experiences with a virtual audience. Furthermore, as part of exposure therapy for individuals with social anxiety disorder [12], patients are exposed to anxiety-evoking situations so that they learn to cope with the anxiety. As it can be very dreadful for some individuals to speak directly in front of a virtual audience, the vicarious experiences could provide them with an opportunity of experiencing the situation with lower level of anxiety. Thus, the first-person-perspective vicarious experience can be used as a first step to increase their motivation to participate in direct experiences.

6

PRESENTATION TRAINING

The study on public speaking found that practicing with a virtual audience is more satisfying and therefore may motivate people to practice. When preparing a speech, people with higher levels of speech anxiety seem often reluctant to rehearse their presentation in front of a real audience [13]. Therefore, getting them to practise would be an essential step forward as practising with audiences improves the presentation [14], reduces anxiety, and increases willingness to speak in public [15]. The results also showed people were more positive towards the effects of practicing with a virtual audience, although no difference was observed in the presentation performance between individuals who practiced with a virtual audience and with an imaginary audience. This finding suggests that trainees might expect greater improvement in their performance after practicing with a virtual audience than after practicing with an imaginary audience, which might result in trainees' overestimation on their own performance. This overestimation however might make the trainees less motivated to practice [2]. To minimize the negative aspects, it might be necessary to make the trainees aware of their own performance, e.g.

by reviewing the audio or video recording of their own presentation practice, and by getting feedback on their performance [16], [17].

6.3. FUTURE WORK

THE research described in this thesis can be extended in many ways. Regarding the expressive virtual audience, the implementation can be improved. For example, more factors could be considered when building the audience's behavioural models, e.g., the relationship between the speaker and the audience, and speech content. The audience could also be more interactive. For example, speech detection could be implemented with the audience so that the audience could be more reactive to the speech. Besides, the audience could also provide feedback on speakers' performance [1] by monitoring this.

This thesis explored the use of practicing with virtual audiences for public speaking training and compared it with practicing with an imaginary audience. However, the effects of practicing with either audience are unclear when comparing it to practicing with a real audience or without any audience. Thus, other control conditions might be included in the future studies to draw further conclusions. Also, only one practicing method was used for each participant. Thus, future studies might investigate the effects of combining practicing methods, e.g., firstly practicing without an audience, then with a virtual audience, and finally with a real audience, to explore an optimal way of improving individuals' public speaking skills.

Concerning the first-person-perspective vicarious experiences, future research should examine the extent to which the vicarious experiences can be effectively used to assist psychological interventions for individuals with mental disorders. For example, individuals with anxiety disorders might profit from this approach by applying the vicarious experience as a first step to increase motivation to participate in direct virtual exposure sessions or exposure in vivo. Moreover, the use of vicarious experiences in skills training could also be of value, such as in helping people to visualize their future performance and master a task in a short time. Furthermore, as the voice of individual's avatar is an effective channel to manipulate the sense of agency in a vicarious experience and thereby influencing the self-efficacy belief, future research could be conducted to examine whether it is also effective in affecting people's attitude and thought.

6.4. TAKE-HOME MESSAGE

THIS thesis aims to explore the effects of public speaking experiences in virtual reality on presenters. To this end, an expressive virtual audience was implemented, which generates flexible expressive behavioral styles by adjusting agent parameters. To provide the foundation for drawing any valid conclusions later on about the effects of the virtual audience, the expressiveness of the virtual audience was validated through a series of perception studies. The results show that people have consistent ideas about audience behavior, and people can recognize different behavioral audience styles. Also, specific postures and behavioral patterns were identified for generic audience styles, which may help designers to develop virtual audiences with noticeable and recognizable behavioral styles. The direct speaking experience with a virtual audience was studied by organizing a public speaking course in an experimental setting. The results showed that individuals seemed to benefit more from a virtual audience than an imaginary audience in reducing speech anxiety. The clearest benefit of practicing with a virtual audience was the satisfaction it gave. Participants were more positive towards training with a virtual audience regarding both the training process and its effect on their presentation ability. A special vicarious experience was created in VR, referred to as a first-person-perspective vicarious experience. The results showed that, when individuals related such vicarious experiences more to themselves, the vicarious experiences was more effective in modifying their self-efficacy. The underlying mechanism that might govern people's self-efficacy beliefs seems closely related to the moderating effect of self-identification on how traditional vicarious experiences affect self-efficacy.

In short, the study demonstrates the possibility of generating virtual audiences with behavioural styles people could recognize. The experience, either indirect or direct, of presenting in front of such an audience can affect the presenter. The effects are important, as well-delivered presentations can have personal benefits, or, in the case of the iconic speeches, change the course of history.

REFERENCES

[1] M. I. Tanveer and E. Lin, "Rhema: A Real-Time In-Situ Intelligent Interface to Help People with Public Speaking," in *ACM international conference on intelligent user interface*, 2015.

[2] D. H. Schunk, "Self-efficacy, motivation, and performance," *Journal of Applied Sport Psychology*, vol. 7, no. 2, pp. 112–137, Sep. 1995.

[3] N. Morina, W.-P. Brinkman, D. Hartanto, I. L. Kampmann, and P. M. G. Emmelkamp, "Social interactions in virtual reality exposure therapy: A proof-of-concept pilot study.," *Technology and health care: official journal of the European Society for Engineering and Medicine*, vol. 23, no. 5, pp. 581–9, Sep. 2015.

[4] S. E. Taylor, T. E. Seeman, N. I. Eisenberger, T. a Kozanian, A. N. Moore, and W. G. Moons, "Effects of a supportive or an unsupportive audience on biological and psychological responses to stress.," *Journal of personality and social psychology*, vol. 98, no. 1, pp. 47–56, Jan. 2010.

[5] M. Chollet, G. Sratou, and A. Shapiro, "An interactive virtual audience platform for public speaking training," in *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, 2014, pp. 1657–1658.

[6] M. Chollet, M. Ochs, and C. Pelachaud, "From Non-verbal Signals Sequence Mining to Bayesian Networks for Interpersonal Attitudes Expression," in *Proceeding of the 14th International Conference on Intelligent Virtual Agents*, 2014, pp. 120–133.

[7] C. Hu, M. A. Walker, M. Neff, and J. E. F. Tree, "storytelling agents with personality and adaptivity," in *Proceeding of the 15th International Conference on Intelligent Virtual Agents*, 2015, vol. 9238, pp. 181–193.

[8] Y. Xu, C. Pelachaud, and S. Marsella, "Compound Gesture Generation[202F?]: A Model Based on Ideational Units," in *Proceeding of the 14th International Conference on Intelligent Virtual Agents*, 2014, pp. 477–491.

[9] N. Kang, "Posture corpus for the behavior generator [Dataset]," 2013. .

[10] O. Kelly, K. Matheson, A. Martinez, Z. Merali, and H. Anisman, "Psychosocial stress evoked by a virtual audience: Relation to neuroendocrine activity," *CyberPsychology & Behavior*, vol. 10, no. 5, pp. 655–62, Oct. 2007.

[11] A. Bandura, *Self-efficacy: The exercise of control*. New York: W .H. Freeman and Company, 1997.

[12] M. Powers and P. Emmelkamp, "Virtual reality exposure therapy for anxiety disorders: A meta-analysis," *Journal of Anxiety Disorders*, vol. 22, no. 3, pp. 561–569, 2008.

[13] J. A. Daly, A. L. Vangelisti, and D. J. Weber, "Speech anxiety affects how people prepare speeches: A protocol analysis of the preparation processes of speakers," *Communication Monographs*, vol. 62, no. 4, pp. 383–397, Jun. 2009.

[14] T. Smith and A. B. Frymier, "Get 'Real': Does Practicing Speeches Before an Audience Improve Performance?," *Communication Quarterly*, vol. 54, no. 1, pp. 111–125, Feb. 2006.

[15] J. Ayres, T. Schliesman, and D. A. Sonandré, "Practice makes perfect but does it help reduce communication apprehension?," *Communication Research Reports*, vol. 15, no. 2, pp. 170–179, Mar. 1998.

[16] M. Arias, P. Pando, A. Rodriguez, P. F. Miaja, A. Vazquez, M. Fernandez, and D. G. Lamar, "The Master's Thesis: An Opportunity for Fostering Presentation Skills," *IEEE Transactions on Education*, vol. 57, no. 1, pp. 61–68, Feb. 2014.

[17] T. Cochrane, "Enhancing oral presentation skills of engineering students: Technology to the rescue with the Virtual-i Presenter (ViP) [online]," in *2009 American Society for Engineering Education Annual Conference and Exposition (ASEE 2009)*, 2009.

Appendix A

The principal component analysis was conducted with varimax orthogonal rotation. Five factors were extracted with eigenvalues over Kaiser's criterion of 1 and in combination explained 74% of the variance. The factor analysis describes how the parameters correlate with the extracted factors. Hence, the parameters with similar factor loadings indicated high correlations between each other, thereby being grouped together (Table A.1). As the factors were independent of each other, the parameter groups were independent of each other if they correlated with different single factors. Thus, there were three independent parameter groups (IG1, IG2, and IG3) and two independent single-parameter groups (IP1 and IP2). Each factor could be interpreted as a characteristic presented by its correlated independent parameters. For example, factor 2 could be interpreted as Patience or Eagerness for information. For the three parameters (DP1, DP2, and DP3) which were correlated with multiple factors, the parameters could also be explained by the correlated factors. For example, the value of Approval (DP2) correlates positively with the values of Eagerness for information (factor 2) and Dominance (factor 3) and negatively with Criticism (factor 4).

Table A.1: Factor loadings of audience parameters and grouping result

Parameters	Loadings on factors					Grouping result
	1	2	3	4	5	
Extraversion	**0.907**	0.036	0.044	0.023	0.093	IG1
Agreeableness	**0.805**	-0.068	-0.002	-0.140	-0.080	
Conscientiousness	**0.670**	0.137	0.277	-0.024	-0.234	
Openness	**0.689**	0.051	0.229	**0.408**	0.307	DP1
Impatience	-0.012	**-0.848**	-0.091	-0.084	0.027	IG2
Eagerness for info	0.090	**0.823**	-0.185	0.133	-0.126	
Interest	0.004	**0.791**	0.276	-0.023	-0.115	
Approval	-0.003	**0.517**	**0.410**	**-0.495**	0.231	DP2
Dominance	0.176	-0.043	**0.856**	0.043	-0.054	IG3
Valence	0.010	0.376	**0.672**	-0.207	0.271	
Neuroticism	**-0.420**	-0.070	**-0.577**	-0.341	0.225	DP3
Criticism	-0.041	0.153	0.032	**0.856**	0.018	IP1
Arousal	-0.030	-0.167	0.000	-0.010	**0.911**	IP2

Note: the loadings larger than 0.4 are in bold type.

Appendix B: Calculating the scale values using the Thurstone's method

The basic method of paired comparisons consists of sequentially presenting pairs of stimuli to an observer and asking the observer which one of the pair has the greatest amount of a certain attribute. Supposing there are n stimuli to compare in total, each observer will have a $n \times n$ matrix of comparison results. If the observer selects stimulus j over i, as having more of the attribute in question, we put a 1 in the j^{th} column and the i^{th} row of a matrix. Using all the matrices of the J observers, a frequency matrix, F, was accumulated. In this matrix, each element is the number of times the stimulus in the j^{th} column was chosen over the stimulus in the j^{th} column. The next step is to form the proportion matrix, P, by dividing each element of F by the number of observers, J. That is, each element of P, i.e., $p_{j>i}$, represents the proportion of observers who select stimulus j over i.

According to Case V of Thurstone model, the scale value difference of two stimuli j and i (i.e., S_j-S_i) can be expressed as the z-score corresponding to the preference frequency (or proportion) of stimulus j over i, $p_{j>i}$. The formula is shown below:

$$S_j - S_i = Z_{j>i} = F^{-1}(p_{j>i}), \tag{B.1}$$

where F^{-1} is the inverse of the standard cumulative normal distribution function.

By transforming each element in the P matrix into a corresponding z-score, a matrix, S, of scale value differences, is then obtained, shown as follows:

$$S = \begin{bmatrix} S_1 - S_1 & S_2 - S_1 & \cdots & S_n - S_1 \\ S_1 - S_2 & S_2 - S_2 & \cdots & S_n - S_2 \\ S_1 - S_3 & S_2 - S_3 & \cdots & S_n - S_3 \\ \vdots & \vdots & \ddots & \vdots \\ S_1 - S_n & S_2 - S_n & \cdots & S_n - S_n \end{bmatrix} \tag{B.2}$$

The scale values of each stimulus can be determined from the column sums of the S matrix. Taking the first column for example, by dividing the column sum by the number of stimuli, we have $\frac{1}{n}\sum_{i=1}^{n}(S_1 - S_i) = S_1 - \bar{S}$. As the average of all the scale values can be set zero, i.e.,$\bar{S} = 0$, the column sums give the scale values directly, i.e., $S_1 - 0 = S_1$.

Appendix C: Descriptions of audience scenarios

Note: Only full scenario descriptions were provided to the participants; the short descriptions are used for convenience when the full descriptions are referred to in this paper.

No.	Short description	Full scenario description
1	Promising business proposal	A person wants to start his own company and needs a sizable amount of investment money for this. He has an opportunity to introduce the investment proposal within 10 minutes to a number of business people, as they will consider whether or not they might invest in this new business opportunity. While listening, the investors find the proposal very promising.
2	Potential business proposal	A person wants to start his own company and needs a sizable amount of investment money for this. He has an opportunity to introduce the investment proposal within 10 minutes to a number of business people, as they will consider whether or not they might invest in this new business opportunity. While listening, the investors find the proposal has some potential but still has some concerns about a number of issues that would require additional work for the person to work out in more detail.
3	Impractical business proposal	A person wants to start his own company and needs a sizable amount of investment money for this. He has an opportunity to introduce the investment proposal within 10 minutes to a number of business people, as they will consider whether or not they might invest in this new business opportunity. While listening, the investors find the proposal definitely impractical.
4	Best man's talk	The best man is talking about an interesting story about the new couple at a wedding party. The people in the party are mostly the new couples' family members and friends. They are friendly and enjoy very much the stories.
5	Friday afternoon lecture	On a Friday afternoon, a teacher is talking about an interesting example in a course. As the content will not appear in the coming exam, the students are more eager to leave.
6	Tuesday morning lecture about exam	On a Tuesday morning, 10 am, a teacher is talking about an interesting example in a course. The content will appear in the coming exam.
7	Attractive Tuesday morning lecture	On a Tuesday morning, 10 am, a teacher is talking about an interesting example in a course. The content attracts the students.
8	Unrelated Monday morning meeting	An employee is presenting his/her work during a meeting with a dozen of colleagues on a Monday morning. The attendees are fellow employees but work on non-related projects. They are indifferent to what the employee is presenting.
9	Related Monday morning meeting	An employee is presenting his/her work during a meeting with a dozen of colleagues on a Monday morning. The attendees are fellow employees working on related projects, so they like to learn how this might affect or benefit their project.

10	Funny show	A beginning stand-up comedian is doing a prepared show for a crowd. The audience is enjoying the jokes and laughing.
11	Not funny show	A beginning stand-up comedian is doing a prepared show for a crowd. The audience is indifferent and finds it not funny at all, and some people are eager to leave.
12	Funeral eulogy	A person is delivering a eulogy at a funeral. The attendees are the family members and best friends of the deceased.
13	Souvenir introduction	At the end of a one-day guided tour, a salesman is introducing a souvenir to the tourists. There is nothing special with the souvenir, and the tourists hope to finish as soon as possible.
14	Positive corporate report	A director of a small company presents a corporate report to all the 11 employees. The report shows improved performance on key targets and they will receive a big annual bonus.
15	Repeated rule announcement	During a weekly school assembly for high school students, the administrator talks about the rules the students should obey, which he or she repeated every week.
16	Announcement of stricter rules	During a weekly school assembly for high school students, the administrator is talking about the new rules the students should obey. The students find the rules much stricter than before.
17	Budget cut	A company manager announces to the work team that a few employees will be made redundant because their budget has been cut. The employees are nervous.
18	Innovative design	At a booth of an exhibition, an exhibitor is introducing a new product. People follow the explanation and find the design innovative.
19	Looking around in an exhibition	At a booth of an exhibition, an exhibitor is introducing a new product. People may only be interested to look around or only to get the freebies.
20	A qualified interviewee	A person is giving a presentation as part of a job interview. The employers find the person sufficiently qualified.
21	A not very satisfactory interviewee	A person is giving a presentation as part of a job interview. The employers find the person may be qualified, but they are not very satisfied with him or her at some points.
22	Training plan	A presentation to fellow sport members at a meeting of the local sport association, about a new idea of this year's training plan.
23	Hobby talk	A student gives a half-hour talk to some fellow students. The student is talking about his hobby which some of the other students also like, but not all.

APPENDIX D

Suppose X_i and Y_i (i = 1, 2, ..., 6) are the parameters for scenario settings X and Y, thereby $X = (X_1, X_2, \cdots, X_6)$ and $Y = (Y_1, Y_2, \cdots, Y_6)$. The observed distance on parameter i ($i = 1,2,\cdots,6$) is noted as d_{Oi}, and the observed distance between setting X and Y is noted as d_O. The expected distance of d_{Oi} and d_O are respectively noted as d_{Ei} and d_E. The observed distance for one parameter, d_{Oi}, is calculated as follows:

$$d_{Oi} = |X_i - Y_i|, \ i = 1,2,\cdots,6. \tag{D.1}$$

The observed distance between two settings, d_O, is the Euclidean distance between X and Y, and can be calculated from the distances for all the six dimensions, i.e., the six parameters listed in Table 3.2:

$$d_O = |X - Y| = \sqrt{\sum_{i=1}^{6} (d_{Oi})^2}. \tag{D.2}$$

To calculate the expected distance between two settings, the expected distances on the six parameters were first calculated. The expected distance for parameter i, d_{Ei} ($i = 1,2,\cdots,6$) is calculated according to the distribution of the possible distances for one parameter. The possible values of d_{Oi} are shown in the Table D.1.

Table D.1: Possible values of the distance for one parameter, $d_{Oi} = |X_i - Y_i|, \ i = 1,2,\cdots,6$

		X_i		
		L	M	H
	L	0	5	10
Y_i	M	5	0	5
	H	10	5	0

Note: the values for L, M, and H were respectively 0, 5, and 10.

As the possible values of a parameter show an equal distribution, i.e., a probability of 1/3 for L, M, and H respectively, the probability for any possible combination of X_i and Y_i is as follows:

$$P(X_i = x, Y_i = y) = 1/3 \times 1/3 = 1/9, x, y \in \{L, M, H\}. \tag{D.3}$$

That is, the probability for any possible distance listed in Table D.1 is 1/9. Thus, the expected distance for one parameter d_{Ei} ($i = 1,2,\cdots,6$) is calculated subsequently:

$$
\begin{aligned}
d_{Ei} = E[|X_i - Y_i|] &= \sum_{x,y \in \{L,M,H\}} |x-y| \cdot P(X_i = x, Y_i = y) \\
&= (0+5+10+5+0+5+10+5+0) \times 1/9 = 40/9.
\end{aligned}
\tag{D.4}
$$

Hence the expected distance between two settings, d_E, is calculated from the distances for all the six dimensions, i.e., the six parameters listed in Table 3.2:

$$
d_E = E[|X - Y|] = \sqrt{\sum_{i=1}^{6} (d_{Ei})^2} = 10.88. \tag{D.5}
$$

Appendix E: An example script of vicarious experience for a female participant

A: audience member; I: the virtual interviewee.
Xxx: participant's name in high-level condition or Jane in low-level condition
H/L Rec: edited recording of the participant's reading in high-level condition or another participant's reading in the low-level condition
Self-introduction: In the high-level condition, the self-introduction of the virtual interviewee was exactly as what the participant had said during the recording in the first session; when in the low-level condition, the city of the virtual interviewee that was mentioned was Delft. The age and occupation was not mentioned.

A-1	Welcome! First, let me introduce the committee members to you. From the window side, they are Russell Fisher, Amy Brown, Sally Wilson, and I am Emma Taylor, the chair. Today you are required to give two short radio lectures on elementary arithmetic. One is on elementary multiplication, and the other is on fractions. After each radio lecture, the committee members will ask you a few questions about that topic. Before you start, would you please introduce yourself briefly?
I	[H/L Rec] My name is Xxx. I'm from ...(place, e.g., country, city). I'm a(n) . . . (occupation) . I'm . . . years old. [Self-introduction] (When in the low-level condition, the city of the virtual interviewee was mentioned as Delft. The age and occupation was not mentioned.)
A-1	Ok, Xxx, you may start now with the lecture on multiplication.
I	[H/L Rec] Let's learn to multiply today. First, let's see what multiplication is. When you multiply, you're adding a number over and over again. Let's first talk about some examples and try to figure out what they mean. If you eat 4 pieces of candy, then you eat another 4, then 4 more, you can say that you multiplied the amount of candy you ate. . . . In summary, multiplication is adding something repeatedly. We learned how to express a multiplication problem. For small numbers, we can either count or add to solve a multiplication problem. Next time we'll learn the times table.

A-1 Thanks for your presentation, Xxx. Now is the question time. Russell, would you like to start?

A-2 Yes. Xxx, since multiplication is just repeated addition, why we still need multiplication?

I [H/L Rec] First, multiplication is a smarter way of calculating the repeated additions. After we learn how to solve multiplications with bigger numbers, such as 25 times 86, you'll find it's much more efficient to use multiplication than adding 86 twenty-five times. Second, multiplication can be used in some other contexts, such as scaling. For example, imagine there're two trees. One is three meters high, and the other is one meter high. Then we could say the big tree is three times taller than the small one.

A-2 Thanks for your answer.

A-3 I have a question. There's a multiplication expression on the cover of my notebook, saying 148 times 210 millimeters. How would you like to explain the meaning of that expression?

I [H/L Rec] It is the size of each page of your notebook. When we learned more about shapes, we'll learn that multiplication can also be used to express the size of certain shapes. The expression is actually 148 millimeters times 210 millimeters. It means the width of your notebook is 148 millimeters, and the length is 210 millimeters.

A-3 Thank you. I'm satisfied.

A-4 Xxx, how would you explain that the order of numbers in a multiplication does not matter, for example, five times seven equals seven times five?

I [H/L Rec] I would start with the interpretation of multiplication problems. For example, there're 5 columns and 7 rows in a table, then how many cells does this table have? To solve this problem, we could calculate either by rows or by columns. That is, there're 7 rows of 5 cells each, so there are 7 times 5 cells in total. We can also look at it in another way. Say that there're 5 columns and each column has 7 cells. That's 5 times 7 cells. Since the number of cells won't change, we know that 7 times5 equals 5 times 7.

A-4 Thank you.

A-1 Xxx, thanks for your presentation. Would you please continue to give a lecture on fractions?

I [H/L Rec] What are fractions? A fraction is part of a whole. It's less than 1 whole thing, but more than 0. We use fractions all the time in real life. Have you ever ordered a quarter-pound burger?

...

Now let's summarize what we've learned today. We learned what fraction is and what it is used for. We know that a fraction is a number; it can be used to represent part of something. Next time, we'll learn how to write and read fractions.

A-1 Thanks for your presentation, Xxx. Now is the question time. Russell, would you like to start?

A-2 Yes. Xxx, as fractions are used to show part of a whole thing, can a fraction be larger than 1?

I [H/L Rec] Yes, later we'll learn that if a fraction is larger than one, we say that it is an improper fraction. For example, we have two pies, and each pie is divided into four parts. If you take one part, how many pies are left? That's seven over four, or seven-fourths. If you have drawn the pies on paper, you'll see that there are actually one and three-fourths of pies left.

A-2 Thanks for your answer.

A-3 I have a question. How would you explain that the denominator can't be zero?

I [H/L Rec] I'll start with the concept of fractions. In a fraction, the denominator represents the total amount, or, the number of parts the whole is divided into, while the numerator represents the amount you have, or, the number of parts you have. The denominator can't be zero because it's impossible to get something out of nothing. So any number over zero is meaningless and undefined.

A-3 Thank you. I'm satisfied.

A-4 Xxx, how would you explain a measuring spoon with a notation of one-fourth teaspoon? It seems strange that a whole spoon is regarded as a fraction, instead of one. Why is that?

I [H/L Rec] The fraction on the spoon represents how much the spoon holds compared with the spoon with a number one on it. That is, if you use the one-fourth teaspoon to fill water into the one teaspoon, you need to do this four times in total. So when you need one-fourth teaspoon of salt, you can directly use that one-fourth spoon. It is also true that we can regard it as one, because we look at it in different perspective. Just as a pizza divided into 8 slices, when you take one slice, we would also call it one-eighth of the pizza.

A-4 Thank you.

A-1 Xxx, I think I can say on behalf of everyone here that your performance today is excellent. Now we can conclude this part of the job interview. Thank you again for your effort.

Appendix F: Questionnaire items for measuring self-identification, presence response, perception of the vicarious experience, and the post-measurement presentation

Questionnaire		Item	label	
			0	10
Self-identification	Q1	I felt as if I was …	watching a video	presenting in front of a panel of professionals
	Q2	I regard the experience as an experience of …	someone else	my own
	Q3	It was easy for me to identify myself with this virtual radio lecture experience.	strongly disagree	strongly agree
	Q4	I felt like being someone else during the virtual radio lecture experience.	strongly disagree	strongly agree
	Q5	I found it easy to relate to the virtual radio lecture experience.	strongly disagree	strongly agree
	Q6	I found it difficult to distinguish myself from the virtual person who gave the radio lecture.	strongly disagree	strongly agree
	Q7	The voice of the interviewee was …	someone else's voice	like my voice
	Q8	The virtual interviewee was …	very different from myself	very similar to myself
Presence response	Q1	To what extent did you find your emotional response and thoughts during the experience similar to those in a real job interview for a radio lecturer?	not at all	very much
	Q2	How often did you find yourself almost automatically behaving within the room as if it was a real job interview for a radio lecturer?	not at all	very much
	Q3	To what extent did you feel as if you were in a real job interview for a radio lecturer?	not at all	very much
Virtual performance	Q1	The radio lecture I experienced was …	very bad	very successful
Capability comparison	Q1	Compared to the radio lecture capability of the virtual interviewee I witnessed, my own radio lecture capability is …	much worse	much better
	Q2	Compared to the radio lecture performance I witnessed, my radio lecture performance would be …	much worse	much better
Presentation performance	Q1	My performance was …	very bad	very successful
	Q2	The panel of professionals were satisfied with my radio lecture.	strongly disagree	strongly agree

Note: all items were rated on a 11-point scale from 0 to 10.

ACKNOWLEDGEMENTS

At this special moment, when I will finish my thesis off, I would like to express my deep sense of gratitude towards my supervisors, colleagues, friends and family. With your guidance, support, and help, being a PhD not only enhances my scientific competence and broadens my view, but also is a most precious experience in my life when I grow and become more mature.

First of all, I would like to thank my promotor Mark Neerincx, and co-promotor Willem-Paul Brinkman and Birna van Riemsdijk. You supported me through my PhD life with your insight, guidance, and patience, without which I would not have completed my PhD research. Besides, special thanks to Willem-Paul Brinkman, who gave me considerable guidance in my research. It is you that showed me the joy of exploring the data which I decided to continue my career on. I am also grateful for your encouragement and understanding when I went through a difficult and exhausting time.

Acknowledgement is also given to Harold Nefs, Judith A. Redi, and Ingrid Heynderickx for sharing their knowledge on experiment and statistical analysis. Besides, I would like to thank our group leader Catholijn Jonker for your support in my research. You created a nice international working environment for us. I also give many thanks to Ruud de Jong and Bart Vastenhouw who help me set up those experiments, and Anita Hoogmoed for your support on all the paper work.

To members and former members of VRET group: Dwi Hartanto, Dylan Schouten, Corine Horsch, Iris Cohen, Wenxin Wang, Yun Ling, Chao Qu, Vanessa Vakili, Myrthe Tielman, Alex Kayal, 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.

Sincere thanks to Rainier R. Kock and Mark J. C. Hendrikx who helped me a lot with my experiments. I also owe a great debt of gratitude to all the II members and friends. There are so many people that I cannot list all your names here. I am grateful that you are always ready to help me with the experiments even though the task were sometimes really boring.

Last but not least, I thank my beloved parents for your encouragement and support, and my husband Junchao for your accompany. To our dearest little boy, Kangning, although you were a so demanding baby that I went through an extremely exhausting time, I still have special thanks as you bring us so much joy and make me stronger, both physically and mentally.

Curriculum Vitæ

Ni Kang

29-04-1985	Born in Benxi, China.

Education

2003–2007	Undergraduate in Measuring Control Technique & Instrumentation Southeast University, China
2007–2010	Graduate in Instrumentation Science & Technology Southeast University, China
2010–2016	PhD. Computer Science Delft University of Technology, the Netherlands *Thesis:* Public Speaking in Virtual Reality: Audience Design and Speaker Experiences *Promotor:* Prof. dr. Mark A. Neerincx

List of Publications

[5] **N. Kang**, W.-P. Brinkman, M.B. van Riemsdijk, M.A. Neerincx, *The design of virtual audiences: Noticeable and recognizable behavioral styles*, Computers in Human Behavior **55**, 680 (2016).

[4] **N. Kang**, W.-P. Brinkman, M.B. van Riemsdijk, M.A. Neerincx, *An Expressive Virtual Audience with Flexible Behavioral Styles*, IEEE Transactions on Affective Computing **4**, 326 (2013).

[3] **N. Kang**, W.-P. Brinkman, M.B. van Riemsdijk, M.A. Neerincx, *Internet-delivered multi-patient virtual reality exposure therapy system for anxiety disorders*, Proceedings of the 29th Annual European Conference on Cognitive Ergonomics (ECCE '11), 233 (2011).

[2] W.-P. Brinkman, D. Hartanto, **N. Kang**, D. de Vliegher, I.L. Kampmann, N. Morina, P.G.M. Emmelkamp, M.A. Neerincx, *A virtual reality dialogue system for the treatment of social phobia*, CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12), 1099 (2012).

[1] D. Hartanto, **N. Kang**, W.-P. Brinkman, I.L. Kampmann, N. Morina, P.G.M. Emmelkamp, M.A. Neerincx,*Automatic Mechanisms for Measuring Subjective Unit of Discomfort*, Stud Health Technol Inform **181**, 192 (2012).