Dissertation thesis

for obtaining the academic degree

Doctor philosophiae (Dr. phil.)

Development and Evaluation of an Interactive Virtual Audience for a Public Speaking Training Application

Submitted to

Institute of Media and Communication Science

TU Ilmenau

by Mariia Dubiago M.A.

1st Reviewer: Univ.-Prof. Dr. phil. habil. Nicola Döring

2nd Reviewer: Univ.-Prof. Dr.-Ing. Alexander Raake

3rd Reviewer: Univ.-Prof. Dr. phil. habil. Nicole Krämer

Submitted on 29.04.2019

Date of the academic debate: 28.11.2019

urn:nbn:de:gbv:ilm1-2019000673

Abstract

Introduction: Fear of public speaking is the most common social fear. Virtual reality (VR) training applications are a promising tool to improve public speaking skills. To be successful, applications should feature a high scenario fidelity. One way to improve it is to implement realistic *speaker-audience interactive behavior*.

Objective: The study aimed to *develop and evaluate a realistic and interactive audience for a VR public speaking training application*. First, an observation study on real speaker-audience interactive behavior patterns was conducted. Second, identified patterns were implemented in the VR application. Finally, an evaluation study identified users' perceptions of the training application.

Observation Study (1): Because of the lack of data on real *speaker-audience interactive behavior*, the first research question to be answered was "*What speaker-audience interaction patterns can be identified in real life?*". A structured, non-participant, overt observation study was conducted. A real audience was video recorded, and content analyzed. The sample resulted in N = 6,484 observed interaction patterns. It was found that speakers, more often than audience members, initiate dialogues and how audience members react to speakers' facial expressions and gestures.

Implementation Study (2): To find efficient ways of *implementing* the results of the observation study in the training application, the second research question was formulated as: "*How can speaker-audience interaction patterns be implemented into the virtual public speaking application?*". The hardware setup comprised a CAVE, Infitec glasses, and ART head tracking. The software was realized with 3D-Excite RTT DeltaGen 12.2. To answer the second research question, several possible technical

solutions were explored systematically, until efficient solutions were found. As a result, self-created audio recognition, Kinect motion recognition, Affectiva facial recognition, and manual question generation were implemented to provide interactive audience behavior in the public speaking training application.

Evaluation Study (3): To find out if implementing interactive behavior patterns met *users' expectations*, the third research question was formulated as "*How does interactivity of a virtual public speaking application affect user experience?*". An experimental, cross-sectional user study was conducted with (N = 57) participants (65% men, 35% women; M_{age} = 25.98, SD = 4.68) who used either an interactive or a non-interactive VR application condition. Results revealed that there was a significant difference in users' perception of the two conditions.

General Conclusions: Speaker-audience interaction patterns that can be observed in real life were incorporated into a VR application that helps people to overcome the fear of public speaking and train their public speaking skills. The findings showed a high relevance of interactivity for VR public speaking applications. Although questions from the audience were still regulated manually, the newly designed audience could interact with the speakers. Thus, the presented VR application is of potential value in helping people to train their public speaking skills. The questions from the audience were still regulated manually by an operator and we conducted the study with participants not suffering from high degrees of public speaking fear. Future work may use more advanced technology, such as speech recognition, 3D-records, or live 3D-streams of an actual person and include participants with high degrees of public speaking fear.

Kurzfassung

Einleitung: Eine der häufigsten sozialen Ängste ist die Angst vor öffentlichem Sprechen. Virtual-Reality- (VR-) Trainingsanwendungen sind ein vielversprechendes Instrument, um die Sprechangst zu reduzieren und die individuellen Sprachfähigkeiten zu verbessern. Grundvoraussetzung hierfür ist die Implementierung eines realistischen und interaktiven Sprecher-Publikum-Verhaltens.

Ziel: Die Studie zielte darauf ab, ein realistisches und interaktives Publikum für eine VR-Anwendung zu entwickeln und zu bewerten, welches für die Trainingsanwendung von öffentlichem Sprechen angewendet wird. Zunächst wurde eine Beobachtungsstudie zu den Verhaltensmustern von Sprecher und Publikum durchgeführt. Anschließend wurden die identifizierten Muster in eine VR-Anwendung implementiert. Die Wahrnehmung der implementierten Interaktionsmuster wurde in einer weiteren Studie aus Sicht der Nutzer evaluiert.

Beobachtungsstudie (1): Aufgrund der nicht ausreichenden Datengrundlage zum realen interaktiven Verhalten zwischen Sprecher und Publikum lautet die erste Forschungsfrage "Welche Sprecher-Publikums-Interaktionsmuster können im realen Umfeld identifiziert werden?". Es wurde eine strukturierte, nicht teilnehmende, offene Beobachtungsstudie durchgeführt. Ein reales Publikum wurde auf Video aufgezeichnet und die Inhalte analysiert. Die Stichprobe ergab $N = 6484$ beobachtete Interaktionsmuster. Es wurde festgestellt, dass Sprecher mehr Dialoge als das Publikum initiieren und wie die Zuschauer auf Gesichtsausdrücke und Gesten der Sprecher reagieren.

Implementierungsstudie (2): Um effiziente Wege zur Implementierung der Ergebnisse der Beobachtungsstudie in die Trainingsanwendung zu finden, wurde die Forschungsfrage wie folgt formuliert: "Wie können Interaktionsmuster zwischen Sprecher und Publikum in eine virtuelle Anwendung implementiert werden?". Das Hardware-Setup bestand aus einer CAVE, Infitec-Brille und einem ART Head-Tracking. Die Software wurde mit 3D-Excite RTT DeltaGen 12.2 realisiert. Zur Beantwortung der zweiten Forschungsfrage wurden mehrere mögliche technische Lösungen systematisch untersucht, bis effiziente Lösungen gefunden wurden. Infolgedessen wurden die selbst erstellte Audioerkennung, die Kinect-Bewegungserkennung, die Affectiva-Gesichtserkennung und die selbst erstellten Fragen implementiert, um das interaktive Verhalten des Publikums in der Trainingsanwendung für öffentliches Sprechen zu realisieren.

Evaluationsstudie (3): Um herauszufinden, ob die Implementierung interaktiver Verhaltensmuster den Erwartungen der Benutzer entsprach, wurde die dritte Forschungsfrage folgendermaßen formuliert: “Wie beeinflusst die Interaktivität einer virtuellen Anwendung für öffentliches Reden die Benutzererfahrung?”. Eine experimentelle Benutzer-Querschnittsstudie wurde mit $N = 57$ Teilnehmerinnen (65% Männer, 35% Frauen; Durchschnittsalter = 25.98, $SD = 4.68$) durchgeführt, die entweder der interaktiven oder nicht-interaktiven VR-Anwendung zugewiesen wurden. Die Ergebnisse zeigten, dass, es einen signifikanten Unterschied in der Wahrnehmung zwischen den beiden Anwendungen gab.

Allgemeine Schlussfolgerungen: Interaktionsmuster zwischen Sprecher und Publikum, die im wirklichen Leben beobachtet werden können, wurden in eine VR-

Anwendung integriert, die Menschen dabei hilft, Angst vor dem öffentlichen Sprechen zu überwinden und ihre öffentlichen Sprechfähigkeiten zu trainieren. Die Ergebnisse zeigten eine hohe Relevanz der VR-Anwendungen für die Simulation öffentlichen Sprechens. Obwohl die Fragen des Publikums manuell gesteuert wurden, konnte das neu gestaltete Publikum mit den Versuchspersonen interagieren. Die vorgestellte VR-Anwendung zeigt daher einen hohen potenziellen Nutzen, Menschen beim Trainieren von Sprechfähigkeiten zu unterstützen. Die Fragen des Publikums wurden immer noch manuell von einem Bediener reguliert und die Studie wurde mit Teilnehmern durchgeführt, die nicht unter einem hohen Grad an Angst vor öffentlichem Sprechen leiden. Bei zukünftigen Studien sollten fortschrittlichere Technologien eingesetzt werden, beispielsweise Spracherkennung, 3D-Aufzeichnungen oder 3D-Livestreams einer realen Person und auch Teilnehmer mit einem hohen Grad an Angst vor öffentlichen Ansprachen beziehungsweise Sprechen in der Öffentlichkeit.

Acknowledgement

This Ph.D. project was possible thanks to several people. First, I would like to thank Prof. Dr. phil. habil. Nicola Doering for granting me the opportunity to conduct the research and being an exceptional example of a successful woman from whom I have learned a lot. I would like to thank Dr. habil. Sandra Poeschl, who supported me and gave a lot of valuable comments, and Dr. Helge Drumm, who introduced the technical side of VR and made the implementation of the VR application possible.

I would like to thank all the organizations that supported, funded, and made the Ph.D. project possible. The research was supported by the Thuringian State Graduate Support, the TU Ilmenau Scholarship, and the Research Group Media Psychology and Media Design. The Competence Center for Virtual Reality at TU Ilmenau provided time in a laboratory that was used for studies. These supports are highly appreciated.

This work would not have been possible without the participation of former students of two Masterclasses who helped in the fieldwork of the evaluation study and voluntarily took part in an observational study.

I was thrilled to be surrounded by such great colleagues. Rohangis Mouhsini, who helped with methodological and statistical insights, with critical comments and encouraging words. Irina Tribussen, Clarissa Treuger, and Anne Hoesch with whom I could have interesting and inspiring professional discussions.

I want to thank my husband, Caglar Ozgur, who supported me in easy and hard moments, inspired me, believed in me, and helped me to keep a positive and motivated attitude. I would like to thank all my friends, who make me happy by being in my life. And I want to thank my family, the most valuable and important thing in my life.

Table of Contents

List of Acronyms

3D	Three-dimensional
CAVE	Cave Automatic Virtual Environment
VR	Virtual Reality
Df	Degrees of freedom
M	Mean
N	Number of cases
SD	Standard deviation
X^2	Chi-square
TV	Television
4k	Four times the pixel resolution
m	Meter
HCI	Human-computer interaction
VRET	Virtual reality exposure therapy

List of Figures

List of Tables

1. Introduction

Public speaking is an essential skill for most professions and in everyday life (Harris, Kemmerling, & North, 2002). Virtual reality public speaking training applications are effective tools to improve public speaking skills. Studies demonstrate that virtual audiences can induce real stress (Pertaub, Slater, & Barker, 2002). Therefore, training with virtual audiences leads to desensitization, resulting in a significant reduction of public speaking anxiety (North, North, & Coble, 1997). To be successful, applications should feature a high scenario fidelity. One way to improve it is to implement realistic speaker-audience interactive behavior. This dissertation thesis aimed to develop and evaluate a realistic and interactive audience for a VR public speaking training application.

First, an observation study on real speaker-audience interactive behavior patterns was conducted (Chapter 2). Second, identified patterns were implemented in the VR application (Chapter 3). Finally, an evaluation study identified users' perceptions of the training application (Chapter 4). Chapter 5 presents a general discussion for all three studies. This describes the relevance of the research (1.1) and chapter overview (1.2).

1.1. Relevance of Research

To be able to hold a speech or presentation in front of an audience is an essential and crucial skill both in private and professional life (at least in many occupations). However, most of the population lacks speaking skills (Hart, Gratch, & Marsella, 2017) and tends to feel stressed before or during public talks. Speaking in public is one of the most commonly reported phobias (Bodie, 2010).

The leading psychological training method for fear of public speaking is cognitive-behavioral training, where people are systematically exposed to fear-triggering stimuli (Heimberg & Becker, 2002; Pribyl, Keaten, & Sakamoto, 2001). As a result of this type of training, fear of public speaking can be reduced, and public speaking skills can be significantly improved (Harris et al., 2002; Pertaub et al., 2002).

For the training of public speaking skills, however, real human audiences are not always available for logistical reasons. Furthermore, a real human audience can be too intimidating for phobic trainees even to start the training. The main alternatives to real human audiences in public speaking training (presentation in front of a real audience, in-vivo exposition) are imagined audiences (imaginal exposure training) and virtual audiences (virtual reality exposure training; Wiederhold et al., 2002).

Wiederhold et al. (2002) claim that virtual reality (VR) exposure training provides more opportunities in comparison to imaginal exposure training. Trainees may not be able to realistically imagine the feared confrontation with an audience, whereas VR technology can simulate the human audience and the public speaking situation quite convincingly. Therefore, VR training leads to a higher level of elicited fear of public speaking than imaginal exposure training (Wiederhold et al., 2002). This is essential since, to change the structure of fear, the feelings must be activated during exposure (Foa & Kozak, 1986). Hence, VR technology has been used in training to create a simulated controllable substitute for a real stimulus (here: the human audience).

VR is a technology rapidly growing in popularity which simulates real or imaginary environments with high realism and interactivity (Bishop & Fuchs, 1992; Zeltzer, 1992). VR has already been actively and successfully used for training

applications since VR training environments allow exposing users to a simulated situation in a safe, highly visual and interactive way (Aldrich, 2004; Ferry et al., 2004). VR training applications, adjusted appropriately, provide a new level of user-system interaction, allowing users to participate actively in virtual situations (Earnshaw, Gigante, & Jones, 1993).

State of research shows that virtual audiences in public speaking applications can induce real stress (Pertaub, Slater, & Barker, 2001) and elicit user responses similar to the ones elicited by real human audiences (Slater, Pertaub, Barker, & Clark, 2006; Zanbaka, Ulinski, Goolkasian, & Hodges, 2007). Further, public speaking training applications are a successful means to reduce public speaking anxiety symptoms and train respective social and presentational skills (Pertaub et al., 2001).

To conduct VR exposure training to improve trainees' public speaking skills, a realistic virtual audience needs to be created. Designing virtual audiences for a public speaking training application requires a deep understanding of how virtual audience is perceived by users and how the perceptions can be influenced by the virtual audiences' design. To be successful, such VR applications should feature users' feelings comparable to feelings in real situations and thus, feature a high scenario fidelity (Bowman & McMahan, 2007a; Kothgassner et al., 2012a). One way to influence the users' perception of scenario fidelity is to implement realistic *speaker-audience interactive behavior*. However, the ways through which humans gain social skills (Slovák & Fitzpatrick, 2015) and data on speaker-audience interaction behavior are scarce (Poeschl, Tudor, & Doering, 2014; Tudor, Mustatea, Poeschl, & Doering, 2014). At the same time, technology can maximize the gain of social skills, but more

interaction between users and the audience should be implemented for higher quality and better outcomes (Poeschl & Doering, 2015; Poeschl & Doering, 2015).

This dissertation thesis aimed to develop and evaluate an interactive audience for a VR public speaking training application and close mentioned gaps in the research. First, an observation study on real speaker-audience interactions was conducted. Second, identified interactive behaviors were implemented in the VR application. Finally, an evaluation study was conducted to identify users' perceptions.

1.2. Chapter Overview

Following the aim of the dissertation thesis, the dissertation is structured into three parts: an observation study on real speaker-audience interactions (Chapter 2); implementation of interactive behavior patterns identified in observation study into VR public speaking training application (Chapter 3). Finally, the evaluation user study on the perceived realism of the new interactive prototype is presented in Chapter 4.

The observation study addresses the behavior of the real audience and real speaker in public speaking situations. Real audience behavior in different situations was observed during different speaker performances to create an empirical basis for realistic virtual audience behavior patterns. In combination with results from the literature review, the most relevant criteria for the implementation of audience behavior was defined. A structured, non-participant, overt observation study on speaker-audience interactive behavior was conducted. An audience with $N = 8$ speakers (4 female, 4 male) and $N = 14$ audience members (9 female, 5 male), all of whom had given informed consent, was recorded using four video cameras. This sample resulted in $N = 6{,}484$ observed interactive behavior patterns that were subjected to quantitative content

analysis. It was found that speakers initiate dialogue by creating questions, and that audience members react to speakers' positive facial expressions and frequency of gestures. Although a convenient international sample with no native English speakers was used, the first study analyzing interactions between a real audience and real speakers was successfully conducted. The results helped to identify interactive behavior patterns for further implementation into VR public speaking training applications.

The implementation study was addressed to find out the best ways of technical implementation of speaker-audience interactive. Based on the observation study, a database of reactions to specific speaker behavior patterns (containing video data of actors) was created, and a suitable algorithm to simulate it was implemented, thereby creating a virtual audience. The hardware setup consisted of a CAVE with three video walls for stereoscopic visualization, Infitec glasses, and ART head tracking. The software setup was realized with 3D-Excite RTT DeltaGen 12.2. Different audio, motion, speech, face recognition was analyzed. As a result of the study, several tools were used for interactive behavior implementation. Manual adjustments were used to manipulate the virtual audience (VA) to ask regarding the presentation content. Self-created audio recognition, Kinect motion recognition, Affectiva facial expression recognition tools were used to manipulate the audience to react to speakers' behavior. Interactive behavior patterns were successfully implemented in the application.

The evaluation study was conducted to find out whether the implementation of speaker-audience interactive behavior patterns met user expectations. A quantitative user study was conducted with (N = 57) participants (65% men, 35% women; M_{age} = 25.98, SD = 4.68). Participants used a public speaking training application (with

interactive or non-interactive options). Four measures of user's perception were assessed: perceived realism, feeling of fear, confidence as a speaker, and feeling of social presence. As expected, a total perception of the feeling of presence and fear in interactive condition was higher than in non-interactive, while the confidence of the speaker in interactive condition was lower than in non-interactive. However, as not expected, perceived realism was lower in interactive conditions. The realism of the new interactive application, at the same time, was evaluated rather high for both conditions. There was a significant difference in user perception of interactive and non-interactive VA.

The dissertation thesis ends with a general discussion that summarizes findings of all three studies, discusses general limitations, and gives opportunities for future research in VR public speaking training application development.

2. Observation Study

The interactive behavior of the people during the communication has been studied a lot. However, the specific case of public speaking application and speaker-audience interaction patterns during this situation were not researched. The study aimed to identify speaker-audience interaction patterns in real public speaking situations.

The following chapter gives the introduction to the observation study (2.1), explains the theoretical background on real speaker-audience interaction (2.2), introduces the methodology of the study (2.3), as well as describes the results of the study (2.4) and summarizes the conclusions (2.5).

The first paragraph (2.1) introduces the topic and relevance of the observation study. The second paragraph (2.2) defines such important for study terms and framework as an audience (2.2.1), speaker-audience interaction (2.2.2), social behavior framework (2.2.3), linguistic communication (2.2.4) including verbal (2.2.4.1) and paralinguistic messages (2.2.4.2) communication, non-linguistic behavior (2.2.5) including gestures (2.2.5.1) and facial expressions (2.2.5.2). The second paragraph is finishing with a summary of all the categories from the social behavior framework that is relevant to the study (2.2.6).

The third paragraph (2.3) introduces research question of the observation study (2.3.1), research design (2.3.2), codebook development (2.3.3), coding method (2.3.4), sample (2.3.5), data collection (2.3.6), coding rules (2.3.7), ethical consideration (2.3.8), and reliability test (2.3.9). The fourth paragraph (2.4) introduces the basic results of the observation study. The results include linguistic verbal communication of real audience

and speaker (2.4.1), paralinguistic messages (2.4.2), gestures (2.4.3), and facial expressions (2.4.4).

2.1.Introduction

To make a professional and successful speech or presentation in front of an audience is an essential and crucial skill for a majority of modern professionals. However, many people tend to feel stressed before or during public talks. These problems are known as public speaking anxiety, stage fright, or fear of public speaking (Kessler, et al., 1994). Consequently, people with such fear are prone to avoid social performance situations (Pribyl et al., 2001).

A number of studies, for example, by North et al. (1997) and by Pertaub et al. (2002), proved that fear of public speaking can be significantly reduced by exposure to the feared situation through a virtual audience. Harris et al. (2002) could show that Virtual Reality Exposure Therapy (VRET) decreases or even erases the fear of public speaking. For the better transfer of gained skills into practice, realistic virtual audiences should be designed for virtual public speaking training applications (Kothgassner et al., 2012a).

One of the leading problems in this area is to reproduce realistic audience behavior, which will be spontaneous and independent from an operator, and realistic audience appearance. Existing VR public speaking applications are missing *realistic interactive speaker-audience behavior*. Speaker-audience interactive behavior is a complicated and complex process. In public speaking situations, the term speaker-audience interaction refers to interaction processes between a speaker and an audience. Every interaction contains linguistic communication (the content of speech and

intonations), and non-linguistic communicative behavior (e.g., gestures, facial expressions; (Hargie, 2006). Both speaker and audience use these two communication channels to interact with each other (Poyatos, 1985).

Designing a virtual audience for a public speaking training application requires a deep understanding of *real human* interactive speaker-audience behavior in real-life public speaking situations. Based on the *social behavior framework* (Hargie, 2006) this research approach views speaker-audience interactive behavior as a social behavior acquiring during interpersonal communications, categorized on linguistic and non-linguistic behavior.

According to the social behavior framework, linguistic behavior includes the actual verbal content and the paralinguistic messages (volume), non-linguistic behavior involves body communication (such as gestures and facial expressions). However, at the moment, there is a lack of detailed data on real human speaker-audience interactive behavior patterns (Poeschl et al., 2014; Tudor et al., 2014). At the same time, without information about real interpersonal communication, "it is difficult to achieve a harmonic and natural man-machine interface for applications such as patient care, geriatric nursing, call centers, psychological consultation, and human communication" (Ren, 2010, p. 7).

Due to the lack of data on real human speaker-audience interaction behavior patterns, the following research question was formulated:

RQ1: What speaker-audience interaction patterns can be identified in real-life?

To answer these research questions, the observation study on real speaker-audience interactions in real public speaking conditions following the categories from

the social behavior framework was conducted. This study analyzes real-life speaker-audience behavior to implement those behavior patterns later into a virtual public speaking training application.

2.2.Theoretical background on real speaker-audience interaction

To understand the nature of speaker-audience interaction, the understanding of the primary communication processes and main communication components is required. First, the general definition of an audience is described as well as its meaning and role, specifically in feared public speaking situations (2.2.1). Later, the definition of the speaker-audience interaction process is given (2.2.2). To describe communication components in detail, the framework combining those components is needed. Thus, the social behavior framework is introduced (2.2.3). All the components of speaker-audience interaction, such as linguistic communication (2.2.4), and non-linguistic communication (2.2.5), are defined.

2.2.1. Audience definition.

Audiences have long been defined as ‘receivers’ of the message presented by a speaker (Adler & Towne, 1996). However, McQuail (1997) argued that although the definition of an audience is clear, it is a complex concept defined by specific factors including: place; people (certain age or gender groups, political belief, or income category); the type of medium or channel involved; the content of its messages; and time. It means that audiences are heterogeneous groups not easily identified for analysis and observation (Moores, 1993).

This research aims to develop a realistic application for public speaking training. Thus, audiences for public speaking trainings are of specific interest to this research. In

public speaking situations, the speaker can be considered as an individual who has the floor, while the audience is considered as people who collected to listen to a speaker (Heylen, 2009). Nevertheless, while training in public speaking skills, people can have a certain degree of fear. Fear of public speaking is a type of social fear, and it is conceptualized as a response to a perceived threat (Beck, Emery, & Greenberg, 2005).

In the social terminology domain, 'audience' is not only used in its typical sense but also refers to “any person or group of people who may potentially perceive an individual's appearance or behavior (including verbal utterances)” (Rapee & Heimberg, 1997). Since the social-evaluated situation is considered as any situation where such an audience exists, the interaction with the audience is not always necessary for fear to occur (Rapee & Heimberg, 1997). The simple appearance of the audience brings the feeling of being observed by another person, and belief that one person is being observed by another (Triplett, 1898). These feelings change the behavior of the individual (Triplett, 1898).

An individual with a fear of public speaking perceives audience behavior differently comparing with people without fear (Rapee & Heimberg, 1997). In particular, people with the fear of public speaking worry less about the environment and factors describing the audience in which they present and focus more on their behaviors and meeting assumed audience expectations (Daly, Vangelisti, & Lawrence, 1989). Low-fear speakers devote more attention to the audience and environment.

Since this study aims to develop a VR training application for both, improving skills and reducing fear, the application should be able to provide different conditions with different level of audience interactivity (e.g., higher interactivity for users with no

fear and willingness to improve skills, and less interactivity for people with high fear). This fact should be considered while designing a realistic audience for training applications.

2.2.2. Speaker-audience interaction.

The public speaking situation is commonly understood as a situation where individuals and an audience having a face-to-face speaking for communication (Poyatos, 1985). Public speaking traditionally takes place in front of a group of people (audience). The content is prepared in advance, and the speech is expected to be presented in a limited time (Lucas, 2009). Thus, in the case of public speaking, the roles in interaction are predefined.

In many books and papers, the process of speaker-audience interaction is simply described as a communication process with a speaker who is actively involved in communication by giving a speech and the listener who is passively involved in the communication process by perceiving and understanding the speech. According to Bakhtin (1999), the ‘listener’ and ‘understander’ (audience) are creating a distorted idea of the real process of the speaker–audience communication. Bakhtin (1999) stated that:

> *“The fact is that when the listener perceives and understands the meaning (the language meaning) of speech, he simultaneously takes an active, responsive attitude toward it. He either agrees or disagrees with it (completely or partially), augments it, applies it, prepares for its execution, and so on. And the listener adopts his responsive attitude for the entire duration of the process of listening and understanding, from the very beginning - sometimes literally from the*

> *speaker's first word. [...] Any understanding is imbued with responsive and necessarily elicits it in one form or another: the listener becomes a speaker." (p. 88).*

Speaker-audience interaction (social behavior) contains verbal (the content of speech), paralinguistic (voice modifications), and non-linguistic (facial and body movements). Both speaker and audience use these three communication channels to interact with each other (Poyatos, 1985). Verbal interaction in public speaking situations may take place at the end of the speech or presentation (Lucas, 2009). However, information exchange between speaker and audience also happens continuously throughout the speakers' speech (Lucas, 2009). The audience, while listening and understanding the meaning of speech, at the same time, take an active and responsive attitude towards the speaker. The audience can either agree, disagree, argue, or so on with the speech. Thus, interpersonal communication is associated with active listening, which means listeners respond to the speaker in such a way that demonstrates interest, understanding, and engagement (Dickson, 2005).

The audience, as a listener, adopts the responsive behavior for the entire duration of speech during listening and understanding the speech (Bakhtin, 1999). Nevertheless, to establish the reaction and response from the audience is a complicated task. Some form of an interpretive framework should be employed to make sense of and to investigate all relevant factors systematically.

2.2.3. Social behavior framework.

People routinely engage in communications with other people, whereby they influence and being influenced by others. Communication between humans is a complex conceptualized as a form of skill (Hargie, 2006). By explaining behavior during communications, people make sense of the social world, they adapt to it, and even shape it. Thus, communication behavior explanations themselves a social behavior, and it must be explained and described.

Social behavior is necessary "to ensure that individuals achieve their desired outcome from social interaction" (Spence, 2003, p. 9). By its definition, social behavior presupposes the involvement of other people and consists of different components. The former is an inherent part of interpersonal contact and include, in order of increasing potency, the mere presence of others, attention from them, and their conversational responsiveness. In terms of overt behaviors during communication, individuals use the response system (voice, hands, face, etc.).

As schematically shown in Figure 1, Hargie's framework categorizes social behavior on linguistic and non-linguistic behavior. Linguistic behavior during interpersonal communication refers to all speech aspects. Linguistic behavior includes the actual verbal content (or verbal communication) and the paralinguistic messages associated with verbal content such as the volume of the speaker. Non-linguistic behavior involves body communication and refers to what individuals do during interpersonal communication. Non-linguistic behavior includes three categories: (1)

Tacesics, the study of body contact, (2) Proxemics, the study of spatial features, and the social distances; (3) Kinesics, the study of body motion (Hargie, 2006).

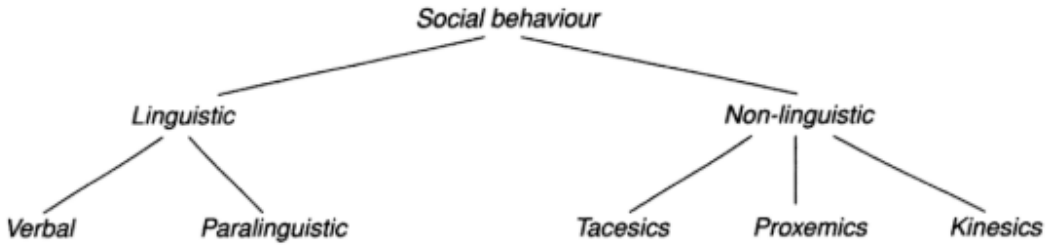

Figure 1. Social behavior framework categories, reported from The Handbook of Communication Skills (p.48), by O. Hargie, 2006, London: Routledge.

Speaker-audience interaction is a very complicated process. However, Hargie's social behavior framework takes into account the central factors of interactions between people that can be applied for public speaking situations. Derailed description and definition of each category will be made in the following paragraphs.

2.2.4. Linguistic communication category.

Linguistic communication is an essential part of human communication. Linguistic communication refers to interactive behavior connected to speech aspects. In particular, linguistic communication refers to actual verbal content (used words and their meanings) of the interaction, and the paralinguistic message that are associated with it. The following paragraphs will describe both types of linguistic communication more detailed.

2.2.4.1.Verbal communication

Humans communicate verbally through speech and language. This allows faster messages sharing, transfer of ideas, and inventions spreading. Verbal communication allows people to communicate messages by representing ideas, objects, places, and others through the language (West & Turner, 2017). Verbal communication defined as a process of sending an encoded message with words from a speaker to a listener (Krauss, 2002) and refers to the use of human language to represent some word and pass of message or information (Ross, 1989). Traditionally, in the case of language, these situations implicate the person or persons to whom the utterance is being addressed (audience in case of public speaking).

Verbal communication consists of some basic components such as the sender (dialogue initiator), message (the type of dialogue), channel, and receiver (dialogue addressees) (Berlo, 1974). Firstly, there should be the initiator of verbal communication. *The initiator of the dialogue* should make an intention to start the communication with another person or number of people. This intention is the result of a stimulus (Fielding, 2006). In the public speaking situation, it can be whether the speaker or audience members.

Secondly, there are different *types of dialogues* (messages) that can be communicated. DeVito (1986) described dialogues messages as “signal or combination of signals that serve as a stimulus for a receiver” (p.201). Different types of messages refer to question, discussion, appeal, feedback, or something else.

Thirdly, it is a channel that described as “vehicle or medium through which signals are sent” (DeVito, 1986, p. 52). In the case of public communication, it will be a speech.

Finally, there is a *dialogue addressee* of the message, “any person or thing that takes in messages” (DeVito, 1986, p. 255). In the public speaking situation, it can be one person or several people. The summary of all the categories relevant to public speaking situations is presented in Table 1.

Table 1

Summary of verbal communication categories relevant for public speaking situations with descriptions and sub-categories

Category	Description and sub-categories
1. Initiation of dialogues	Intention to start the communication with another person or number of people: • Speaker • Audience member
2. Types of dialogues (messages)	Types of signal or combination of signals that serve as a stimulus for a receiver: • Questions • Answers
3. Addressees in dialogue	Any person or thing that takes in messages • One-to-one • One-to-many

2.2.4.2.Paralinguistic messages

At the same time, communication between people is not only what humans say, but it is also about how they say it. Linguistic communication incorporates all components of speech, not only actual words used (verbal communication) but features of their delivery paralinguistic messages. Paralinguistic messages referred to "the vocal characteristics of the verbal behavior" (Hargie, 2006, p. 350). The featured of paralinguistic commanding the attention are intonations, the volume of voice, response latency, duration, and fluency. Since in public speaking situations, the speaker is most of the time holding the monologue, duration of speeches are predefined, speech is prepared in advance (thus fluency is affected), and there is no verbal response. Thus, such features can be count as irrelevant for this research. Intermediate levels of volume and intonation appear to characterize effective communication assertion.

Intonations are considered as one of the most critical features of an effective statement (Romano & Bellack, 1980). Intonations refer to acoustic parameter combinations to communicate discourse meaning (Levis, 2013). Voice intonations are responsible for about 38% of the message perception, while an actual word for only seven percent (Chibelushi & Bourel, 2003). In social behavior, intonations help to deliver the correct message regardless of the meaning of the word. Intonations are used to carry different kinds of information and refer to the means of information delivery during a speech, which is sometimes independent from the words meanings (Nolan, 2006). Traditionally, it has been claimed that there are emotion-specific intonation patterns are existed (Fonagy & Magdics, 1963).

Intonations signal not only a grammatical structure but also such information as indicating that the speaker has said everything that he/she wanted to say show that speaker is inflow and should not be interrupted, as well as that the speaker requests for a respond or feedback from the listener (Nolan, 2006). Specific rise and fall of the pitch on certain words place a critical role in delivering the correct message. Intonation adds a new layer of meaning to verbal content, and it helps speakers to communicate through the fall and rise of the voice (UTS: HELPS, 2009). There are five main kinds of intonations that can be recognized, using the same words. These five intonations can be used to deliver five different kinds of messages (UTS: HELPS, 2009). The intonations recognized by the rate of speech, by the rise of voice, and by the and fall of the voice. Intonations add another meaning to the verbal content. There are five main types of intonations which can deliver a different kind of messages. These types are briefly described below:

1. Question: this intonation indicates that the question is being asked. Question intonation is most of the time marked in speeches by rising pitch when the tone goes from low-to-high towards the end of the statement. The end is slightly prolonged as compared to the rest of the statement. The grammatical meaning comprises of different tones, which indicate the types of questions asked by the speaker. Since not all questions such as what, when, why, where, who, and how require the tone to rise and fall, intonations differentiate the intended meanings. Even though sometimes the grammar indicates a statement.
2. Complete or incomplete statements: intonation that informs the listener whether the speaker has finished the statement or not. Usually, the speaker pauses before

starting a new statement. Nevertheless, sometimes the statement remains incomplete, so the speaker takes a pause and starts another statement instead of finishing the previous one. The tone throughout the incomplete intonation remains the same, the final syllable is not prolonged, and the ending is abrupt and unexpected.

3. Doubtful statement: this intonation deals with doubtfulness of the communicated information. During this intonation, the speaker pauses before the start, and the tone at the beginning of the statement moves gradually from a high-to-low, then, from the middle of the statements towards the end. The tone stays high. The final syllable is slightly prolonged with the increase in pitch, and the end of the statement is abrupt and unexpected.
4. Confident statement: the intonation shows the certainty and confidence about the information which is being delivered. Speaker takes a pause before the beginning of a statement. The tone of the statement moves from high-to-low from the middle towards the end. The final syllable of the statement is not prolonged, and the end is not abrupt.
5. Emphasizing statement: the meaning is conveyed by the intonation pointing towards how open or friendly the speaker is towards the listeners. Speaker takes a pause before the beginning of the statement. The tone moves from low-to-high moving from the middle towards the end of the statement. The final syllable is not prolonged. The end is not abrupt (UTS: HELPS, 2009).

Delivering the intended meaning through intonation is not always an easy thing to do and can lead to irritation, confusion, or misunderstanding of the listener. The

listener might not be able to distinguish the intonation information, completeness or incompleteness of statement, interaction, or response ques. If the speaker lacks confidence, it can influence his/her voice and intonations, which could decrease the understanding between speaker and audience even more (UTS: HELPS, 2009).

Another paralinguistic message used by the speaker is *volume*. There is no formal definition that exists at the moment for what the volume of sound means. However, the volume is generally used as a synonym for loudness (Oland & Dannenberg, 2017). American National Standards Institute defined volume or loudness as "...that attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from quiet to loud." (American National Standards Institute, 1973). Volume is a perceptual concept that states that the human ear's sensitivity to sound varies on frequency. The data on the volume, in contrast to the previous two paralinguistic messages, are consistent: "effective conflict assertion is characterized by an appropriate, moderate volume that is louder than the speech produced in ordinary conversation (e.g., Rose & Tryon, 2016) and by non-assertive persons (e.g., Eisler, Miller, & Hersen, 1973)." (Hargie, 2006, p. 355).

All the categories for the paralinguistic messages relevant to public speaking situations are summarized in Table 2.

Table 2

Summary of paralinguistic messages categories relevant for public speaking situations with descriptions

Paralinguistic Messages Category	Descriptions
1. Question	• The tone moves from low-to-high towards the end of the statement. • The final syllable is slightly prolonged. • The end is slightly prolonged.
2. Incomplete statement	• Speaker takes a pause before the beginning of a statement and no pauses in between. • The tone remains the same throughout the statement. • The final syllable is not prolonged. • The ending is abrupt and unexpected.
3. Doubtful statement	• The speaker takes a pause before the statement. • The tone moves from high-to-low at the beginning of the statement. • The tone is high from the middle towards the end of the statement. • The final syllable is slightly prolonged with the increase in pitch. • The end of the statement is abrupt and unexpected.

Table 2

Summary of paralinguistic messages categories relevant for public speaking situations with descriptions

Paralinguistic Messages Category	Descriptions
4. Confident statement	• Speaker takes a pause before the beginning of the statement. • The tone moves from high-to-low moving from the middle towards the end of the statement. • The final syllable is not prolonged. • The end is not abrupt.
5. Emphasized statement	• Speaker takes a pause before the beginning of the statement. • The tone moves from low-to-high moving from the middle towards the end of the statement. • The final syllable is not prolonged. • The end is not abrupt
6. Volume	Subjectively speakers' volume level while delivering the speech.

2.2.5. Non-linguistic communication.

Non-linguistic communication conveys a large amount of information in an interaction (Mcfall, Winnett, Bordewick, & Bornstein, 2016), as well as in general, in interpersonal communication. According to the social behavior framework, non-linguistic behavior consists of tacesics, proxemics, and kinesics. However, in the case of a public speaking situation, tacesics and proxemics can be considered irrelevant.

Tacesics is the study of body contact, which is not possible between speaker and audience, while proxemics is the study of spatial features and the social distances which are perfidies and cannot be changed during the public speaking. At the same time, researchers have examined the contribution of gestures, facial expressions, eye contact, and gestures make to effective conflict assertion.

2.2.5.1.Gestures

Most of the children use gestures before speaking (Goldin-Meadow & Alibali, 2013). Gestures were used as communication between people in the early stage of humans. These are the first evidence of gesture significance in interaction. At the same time, the recipient's (e.g. audiences') reaction towards the gestures is deemed necessary to interpret (McNeill, 1995).

In the literature, there is an abundance of gestural typologies, but there is an agreement among researchers regarding what distinctions are necessary or useful (Krauss, Chen, & Chawla, 1996). Krauss et al. (1996) stated that "all gestures are hand movements, but not all hand movements are gestures […]". As Hummels and Stappers (1998) explain "a gesture is a movement of one's body that conveys meaning to oneself or a partner in communication". Abner et al. (2015) gave a more recent definition of gestures: "spontaneous movements of the hands and body that universally accompany speech" (p. 437). There are two main types of gestures used during the speech: communicative – gestures that accompany, intensify, modify or support the speech and

informative – gestures that provide information about the speaker as a being (Ekman & Friesen, 1969).

Speaker's communicative gestures are divided into manual gestures and non-manual gestures. Manual gestures were categorized into symbolic, deictic, and beat gestures. Symbolic gestures described as hand movements with widely-recognized conventionalized meanings (Ricci Bitti & Poggi, 1991). Symbolic gestures refer to hand movements that can be used as word replacement, with widely-recognized conventionalized meanings (Ricci Bitti & Poggi, 1991).

Deictic gestures consist of pointing or indicative movements. They are typically formed with the index finger extended, and the remaining fingers closed (Abner et al., 2015). Deictic gestures used to indicate persons, locations, directions, or objects, as well as to "point to" imaginary, abstract, or unseen things (Abner et al., 2015). Deictic gestures function in a similar way to demonstrative pronouns, such as 'this' and 'that' (Abner et al., 2015). They can accompany speech or can also be used as a substitute for demonstrative pronouns.

Beat gestures referred to as simple, repetitive, rhythmic movements that bear no apparent relation to the semantic content of the accompanying speech (Feyereisen, Van de Wiele, & Dubois, 1988). The beat gestures refer to motor gestures, and can also be called as "batons" (Efron, 1972) and "beats" (Kendon, 2002; McNeill, 1995). Bull and Connelly (1985) stated that motor gestures are coordinated with the speech. Motor gestures during the speech tend to fall on stressed syllables (McClave, 1994). However, the synchrony is far from perfect.

Elliott et al. (2004) categorized head movements, shoulder movements, body movements (including body postures), and non-manual gestures. As stated by Kendon (2002) the head shakes can be used in many different contexts and can be interpreted as negation. At the same time, the head shakes are also used by speakers to comment on his/her utterance (self-corrections, expressing doubt about what he/she is saying, expressing doubt about whether he/she had responded to interlocutor appropriately; Kendon, 2002).

Shoulder movements relate to the speaker being confident or not confident about what to say (Jokinen & Allwood, 2010). Body movements can be identified by two categories: moving forward – the whole trunk forward movement; move backward –the whole trunk backward movement (Allwood & Cerrato, 2003). There are limited studies that define and interpret these body movements. One of the interpretations of leaning front and back is made by Siegman and Feldstein (1987) in Nonverbal Behavior and Communication stated that body movement as lean forward, occurs more among people who sympathize with each other.

The body posture of the speaker is manlily focused on open-closed positions of arms or legs (Siegman & Feldstein, 1987). Open posture (arms are apart and knees separated) shows confidence while closed posture (arms are closed or folded together and legs crossed) displays nervousness (Navarro & Karlins, 2008, 2007).

Speaker's gestures attract audiences' attention. Broaders and Goldin-Meadow (2010) elaborate that the speaker's gestures during his/her talk do not go unnoticed by the listeners. The attention that has been decreased throughout the speech can be brought back by gesture use. Listeners can successfully comprehend speech through the

speaker’s gestures. Nevertheless, if the speech is difficult to understand, either because it is difficult, ambiguous, or unclear for the listeners’ skills, a gesture can provide a second information channel that increases the probability of successful comprehension (Goldin-Meadow & Alibali, 2013). Some studies support the statement that gestures affect the audience's understanding of the speech. Hand gestures are congruent with the verbal content of the speaker’s message can enhance comprehension of the message (Young, 2017). Gestures that express information that is not expressed in speech can distract listeners’ direct uptake of the information in speech (Goldin-Meadow & Sandhofer, 1999).

Speaker’s gestures provide for the audience additional information to the main speech content. In the service of constructing meaning, listeners recruit information from their perceptual-motor system (Goldin-Meadow & Alibali, 2013). The sensitivity of the audience towards the speaker’s gestures plays an essential role in speaker-audience interaction; either it leads to successful or failed interaction. The summary of all categories and sub-categories relevant for public speaking situations with descriptions presented in Table 3.

Table 3

Summary of gestures' categories and sub-categories relevant for public speaking situations with descriptions

Gestures Category	Sub-Category	Description
Symbolic	• Handwave/salute • Thumbs up / down • Clenched fist • Crossed fingers • Clap • Hand rubbing • Praying hands • Showing numbers with fingers • OK-sign	Hand configurations and movements that can be used as a replacement for words, with recognized conventionalized meanings.
Deictic	• Pointing, indicating at a person • Pointing, indicating at direction or object	Indicative or pointing movements, typically formed with the index finger extended and the remaining fingers closed.
Motor	• No beat • Beat	Simple, repetitive, rhythmic movements that bear no obvious relation to the semantic content of the accompanying speech.

Table 3

Summary of gestures' categories and sub-categories relevant for public speaking situations with descriptions

Gestures Category	Sub-Category	Description
Head movements	• No movement • Head nod • Head shake	Forward movement of the head going up and down, which can be multiple (vertical). Left-right or right-left movement of the head which can be multiple (horizontal).
Shoulder movement	• No shrug • Shoulder shrug	Lifting of both shoulders, full: up and down again.
Body movements	• No movement • Moving of trunk (forwards, backward or sideways	The person is standing still or moving very little. Clear forward, backward, sideward movement of the whole trunk.
Body posture	• Posture open • Posture closed	Hand and arms are apart and knees are separated. Arms are closed or folded and legs crossed for closed (building a "circle").

2.2.5.2.Facial expressions

Ekman (1977) analyzed and synthesized the experiments during half of the century and confirmed that "accurate judgments of facial expression can be made" (p. 98). Similarly, the research by Oh et al. (2016) points out that "people tend to rely on

these nonverbal behaviors to detect deception and/or form interpersonal judgments" (p. 3).

From the perspectives of the speaker-audience interaction in a public speaking environment, facial expressions are considered as audience members' feedback forms and as a speakers' message form. Basic expressions of emotion via facial mimics are universal, Ekman and Friesen (1971) reported six facial expressions through the face that are main and rapidly recognized across different cultures: anger, happiness, fear, surprise, disgust, and sadness.

The Facial Affect Scoring Technique, FAST (Ekman, Friesen, & Tomkins, 1971), is one of the observer-based systems for the facial expression measurements. The system uses a series of pictures of the lower face, of bows-forehead, eyes, and the whole face to illustrate each of the six emotions. Based on the facial muscle movement analysis, a more comprehensive and objective Facial Action Coding System (FACS) was later developed. FACS can be a tool for measuring emotion-related facial expressions as well as recognizing visible facial expressions (Ekman, 1977; Ekman & Friesen, 1976; Ekman, Friesen, & Hager, 1978). Therefore, FACS can be the basic framework for identifying any visible facial movements based on minimal action units (Ekman, 1977).

Action units refer to the smallest facial movements. FACS system specifies nine action units in the upper part of the face, 18 units in the lower part of the face, 14 units in the head movements and positions, nine in the eye movements and positions, five different action units, nine action descriptors units, nine gross behaviors units, and five visibility codes. Each action unit has a verbal and numeric label and a specific

anatomical basis in facial muscles (Cohn, Ambadar, & Ekman, 2007). FACS can be used by coders to decompose and summarize visible facial expressions of the speakers and audience members in the videos in specific time units.

Six facial expressions were included in the study: eye contact, neutral, positive, negative, surprised, and unclassified. Eye contact is a direction where the individuals are looking at. In public speaking situations, audience members can look at the speaker, at some objects, or another audience member.

The second facial expression is neutral, and it is described as a face that is not showing any action. The eyes during the neutral facial expression are open, the jaw is closed, and no action units can be recognized.

The third facial expression is positive facial expression. It can be recognized with the lip corner puller (smile) and cheek riser.

The fourth facial expression is negative. It can be recognized when the inner brow raiser, brow lowered, and lip corner depressor action units are coded on the face.

The fifth facial expression is a surprised one. Surprising facial expression can be recognized by such action units as raised eyebrows, raised the upper lid of eyes, widened eyes, dropped open jaw, relaxed lips, and no activity in the neck.

Finally, there is an unclassified facial expression category. In public speaking situations, there are cases when the audience member might not be visible: covered with the audience member from the front row, by some objects in hands, or turned into a direction that is making a face invisible from some angle. All the categories for facial expressions relevant to public speaking situations are summarized in Table 4.

Table 4

Summary of facial expressions' categories relevant for public speaking situations with descriptions

Facial expression category	Description (action units)
Eye contact	At speaker
	At object
	At another audience member
Neutral	The face is not showing any action
	Eyes are open
	Jaw is closed
	No action units
Positive	Lip corner puller
	Cheek raiser
Negative	Inner brow raiser,
	Brow lowerer,
	Lip corner depressor
Surprised	Eyebrows raised,
	The upper lid of eyes raised, eyes widened
	Dropped open jaw, lips relaxed
	No activity in the neck
Unclassified	Facial expression is not recognized
	Action unit is not recognized or not visible

2.2.6. Summary.

Following the social behavior framework, there are two types of social behavior that can describe speaker-audience interactive behavior: linguistic and non-linguistic. Linguistic behavior divided into verbal and paralinguistic behavior. Verbal behavior includes such sub-categories as dialogue initiator, dialogues addressees, and types of speech. The paralinguistic category involves seven sub-categories: question intonation, incomplete statement intonation, doubtful statement intonation, confident statement intonation, emphasizing statement intonation, volume.

Non-linguistic social behavior divided into gestures and facial expressions. Gestures include four sub-categories: symbolic, deictic, beat, and non-manual. Finally, facial expressions consist of six sub-categories: eye contact, neutral, positive, negative, surprised, and unclassified facial expressions.

2.3. Methodology

The goal of the study presented here was to reveal what patterns of speaker-audience interaction behavior can be identified in real-life to transfer these behaviors later to virtual application. To acquire data about real speaker-audience interactive behavior patterns, an observation study was conducted (Vaus, 2001). This method has been already used earlier in studies for similar purposes (e.g., Poeschl & Doering, 2012b). This method is fitting the study and purpose of the study the most because it allows quantification of the observed behavior patterns data.

The following paragraphs present the research question of the observation study (2.3.1), research design (2.3.2), codebook development (2.3.3), coding method (2.3.4),

sample (2.3.5), data collection (2.3.6), coding rules (2.3.7), ethical consideration (2.3.8), and reliability test (2.3.9).

2.3.1. Research question.

The observation study focused on real speaker - audience interactive behavior patterns. The research question of the observation study is:

RQ1: What patterns of speaker-audience interactions can be identified in real-life?

Speaker-audience interactive behavior patterns were coded, and the relationship between a speaker's behavior and audience members was analyzed. The answer to this question will help to implement interactive patterns into a virtual public speaking application. These criteria of speaker-audience interactive behaviors are explained in the codebook development section.

2.3.2. Research design.

To answer the research question, the descriptive research was used. The structured, non-participant overt observation study on speaker-audience interactions during real presentations and their subsequent discussions in an English-language master program class at a middle-size university in Germany were conducted. The observation method was used because it allows a thorough quantification of the observed behaviors.

Following the aim of the research (to identify the interactive behavior patterns between real audience members and real speakers), the observation of behavior was done in a natural setting, without any attempt to intervene. Non-participant observation approach was chosen. The research was interested in reified patterns that emerge from

public speaking situations (Given, 2008). Speakers, as well as audiences, were video-recorded with four video cameras to capture different angles.

2.3.3. Codebook development.

Quantitative content analysis research method followed to analyze the videos empirically. This method is widely used in social sciences to investigate the object and social issues empirically. The usability of this method ranges from written text, either print or digital, to videos, films, and images. Berelson (1984) explained quantitively content analysis as a research process for describing manifest communication content in an organized, objective, and quantitative manner. Riffe et al. (2005, p. 5) elaborate that:

> *"Quantitative content analysis is the systematic and replicable examination of symbols of communication, which have been assigned numeric values according to valid measurement rules and the analysis of relationship involving those values using a statistical method, to describe the communication, draw inferences about its meaning, or infer from the communication to its context, both production and consumption."*

This study is conducted as a didactic reason to demonstrate linguistic and non-linguistic interactive behavior patterns that have been used in a public speaking situation by both: the speakers and audience. The quantitative approach has an advantage for the study by reducing the large sets of data and to characterize the data with summary statistics such as ranges, averages, and percentages (Riffe et al., 2005). This study uses a quantitative research method to find out the relationship between different variables to describe the correlation the two have within a population. Quantitative research also gives a structured measure (Hopkins, 2008). This section discusses the coding methods

which have already been discussed in the literature. The section explains how the existed methods have been implemented to develop the codebook

2.3.4. Coding method.

In social and personality psychology, the observation of behavior is placed in the center of all scientific research. Although there are a variety of different methods that researchers use while observing behavior, the term "behavioral observation" usually refers to a researcher seeing and/or hearing, and later systematically recording, the individuals' or groups of individuals' behaviors within a particular context of interest (Heyman, Lorber, Mark Eddy, & West, 2014).

There are two main coding methods to identify categories for behavior patterns. The first method codes movement features such as body actions and postures (Dael, Mortillaro, & Scherer, 2012). To code emotion expressions, the coding system contains $N = 141$ behavior variables—the second method codes movement types (Dael et al., 2012). To create behavioral combinations for virtual humans, Kang et al. (2013) used this coding method to identify the head, gaze, arms and hands, torso, and leg postures within a real audience.

The codebook included detailed positions of hands and movement directions. However, it only provided behavior combinations and did not focus on identifying interactive behavior patterns. Bakeman and Gottman (1997) stated that borrowing coding schemes from other researches rarely is appropriate since the development of codes is very much dependent on theory: Codes should represent the specific problem within the study. The second method uses variable construction based on theory (Harrigan, Rosenthal, & Scherer, 2008). It adapts body movements typology to the

purpose of the study to address specific categories relevant to a research aim (Dael et al., 2012).

This observation study combines both coding methods. Non-linguistic behavior was coded categorically. All the categories were adapted to the study purpose. The features of the eye gaze, intonations, dialogues categories were simplified according to the study purpose and simplify the coding process. Four main categories of interactive behavior patterns were extracted following the social behavior framework. Two categories were developed to code the linguistic behavior: verbal content in speaker-audience dialogues and speakers' intonations. Two categories were developed to code non-linguistic behavior: gestures and facial expressions (including eye contact).

There are two methods to determine and code the behavior categories. The first one is an event sampling, also called as frequency counts. It involves observation of targeted behavior categories or specific events that has been predefined in advance (e.g., dialogues, only when they occur between the speaker and audience members). Event sampling is generally used to determine the frequency (or how often) a specified behavior or specific event occurs (Hills, 1993).

The second method is a time sampling method. The time sampling method is used to observe behaviors or events that occur quite frequently for efficient event sampling and determines the degree of how much a behavior or an event occurs by making observations at the specific time intervals (Hills, 1993). Basically, during predetermined periods, as in this study, the first five seconds of every 30 seconds of presentation, the coder observes and codes every occurrence of behaviors or events. In this observational study, both methods were combined. Event sampling was used to

code such rare used events as paralinguistic messages and verbal dialogues, while time sampling was used to code such frequently used non-linguistic events as facial expressions, eye-contact, and gestures.

Verbal communication. According to Miles and Huberman (1994), codes are termed as labels or tags for associating meaning to expected or disruptive information arranged together during the study. Codebook comprised of code sets, coding instructions, and examples to make it more consistent. The part of the codebook regarding verbal communication contains components based on Hargie (2006) social behavior theory and on Berlo (1974) categories: who initiated the dialogue (sender), addressees in dialogues (receiver), types of the dialogues (messages). Such a category as the channel was not included in the codebook since the only verbal channel was relevant for study and was annualized. An inductive approach was followed to develop the categories by carefully examining raw data. The development of codebook is a repetitive process and requires revisions unless required reliability is achieved.

Paralinguistic messages. The codebook was developed inductively through analyzing research done previously on the topic of intonations, and deductively adjustments were made during the study and pre-test. Speaker codebook adapts to the definition of intonation given by UTS: HELPS (2009) that there are five types of intonations or five types of meanings that can be conveyed using intonations. These intonations include a confident statement, questionnaire statement, incomplete statement, doubtful statement, and emphasizing statement.

Gestures. Inductive and deductive methods were used for developing the codebook. The codebook was divided into two parts: gestures made by the speaker

while presenting in front of the audience, and gestures of audience members made while listening to the speaker`s presentation. Some categories were used for both parts of the codebook; however, some categories were only suitable only for one part.

Speaker communicative gestures were divided into manual gestures and non-manual gestures. Manual gestures were categorized into symbolic, deictic, and beat gestures. Based on Ricci Bitti and Poggi's (1991) codebook, included symbolic gestures as handwave/salute, thumbs up / down, clenched fist, crossed fingers, clap, hand rubbing, praying hands (see the example on Figure 2), showing numbers with fingers, OK-sign.

Figure 2. Example of symbolic gesture performed by speaker. Speaker is placing hands in a praying hands position.

Deictic gestures referred to indicative or pointing movements, typically formed with the index finger extended, and the remaining fingers closed (Abner et al., 2015). The codebook included two types of deictic gestures: pointing with the finger to a person (Figure 3) and pointing with the finger to a direction or an object.

Figure 3. Example of deictic gesture performed by speaker. Speaker is pointing out the audience member whose question he wants to hear.

Beat gestures referred to as simple, repetitive, rhythmic movements that bear no apparent relation to the semantic content of the accompanying speech (Feyereisen et al., 1988). The codebook involved the presence of beats.

Non-manual gestures Elliott, Glauert, and Kennaway (2004) categorized head movements, body movements, and shoulder movements as non-manual gestures. It can be concluded that non-manual gestures involve all gestures that do not involve hands. Codebook involved four categories as head movements, shoulder movements, body movements, and body posture.

Head movements referred to the head forward movement going up and down (vertical), which can be multiple and left-right or right-left movements of the head (horizontal), which can be multiple (Allwood & Cerrato, 2003). Shoulder movement referred to the lifting of both shoulders, full: up and down again (Jokinen & Allwood, 2010). The codebook included the presence of such a shoulder movement. Body movements referred to as clear forward, sideward, or backward movement of the whole

trunk (Allwood & Cerrato, 2003), and the codebook included the presence of such a movement. Finally, body posture referred to open and closed posture, where open posture is the posture where hand and arms are apart, and knees are separated, and closed posture is the posture where arms are closed or folded and legs crossed for closed (building a "circle") (Siegman & Feldstein, 1987). Lexical gestures, another category of manual gestures, were not included in the research and coding since it is a highly disputed definition and also due to its irrelevance for the research.

Audience gestures were divided into four categories by using deductive and inductive methods: head movements, body movements, body posture, and hand movements.

The head movements category defined whether the audience member is moving his/her head (shaking or nodding) or if he/she has a stable head position. The body movements category identified if the audience member expressed no body movements or were moving the whole trunk sideward /forward /backward (Allwood & Cerrato, 2003). The body posture category defined which of the posture audience member employed: closed or open posture. The closed posture was categorized as the posture when the audience member's arms are folded in front of the chest, and open posture was categorized as the posture when the audience member's arms and shoulders are relaxed (not folded) (Siegman & Feldstein, 1987). The hand movements category defined if audience members had a stable hand position, raised his/her hand to ask a question from the speaker or give a comment for the speaker's speech, or if audience members used adapters – of some object (e.g., clothing, pencils, smartphones) or manipulating either of the person (scratching, tapping oneself) (Ekman & Friesen, 1969). All the categories and

sub-categories describing social behavior are summarized in Table 5 and were used for coding speaker-audience interactive behavior.

Table 5

Summary of types of social behavior, categories, and sub-categories following the social behavior theory

Type of social behavior	Categories	Sub-categories
Linguistic	Verbal	1. Dialogue initiator (Speaker, Audience).
		2. Addresser in the dialogue (One-to-One; One-to-Many).
		3. Types of Speech (Questions, Answers).
		4. The initiator of the dialogue.
	Paralinguistic	1. Question intonation.
		2. Incomplete statement intonation.
		3. Doubtful statement intonation.
		4. Confident statement intonation.
		5. Emphasizing statement intonation.
		6. Volume.

Table 5

Summary of types of social behavior, categories, and sub-categories following the social behavior theory

Type of social behavior	Categories	Sub-categories
Non-linguistic	Gestures	1. Symbolic (handwave/salute, thumbs up / down, clenched fist, crossed fingers, clap, hand rubbing, praying hands). 2. Deictic (pointing with the finger to a person or object). 3. Beat. 4. Non-manual (head movements, shoulder movements, body movements, and body posture).
	Facial expressions	1. Eye contact (at the speaker, at an object, at another audience member). 2. Neutral. 3. Positive. 4. Negative. 5. Surprised. 6. Unclassified.

2.3.5. Sample.

Firstly, a priori power analysis was conducted using G*Power3.1 (Faul, Erdfelder, Lang, & Buchner, 2007). There was no prior analysis performed for verbal communication since the date was only analyzed with descriptive tests. For paralinguistic messages (with 4 groups of different intonations) with a medium effect size (d = 0.3), and an alpha of 0.05 for achieving a power of .80, a total sample of 122

coded intonations were required. For gestures (with 2 groups of using or not using gestures) with a medium effect size (d = 0.3), and an alpha of 0.05 for achieving a power of .80, a total sample of 88 coded gestures were required. Finally, for facial expressions (with 3 groups of different facial expressions) with a medium effect size (d = 0.3), and an alpha of 0.05 for achieving a power of .80, a total sample of 108 coded facial expressions were required.

The non-participant observation was conducted. N = 13 speakers (8 female and 5 male) and N = 17 (9 female and 5 male) audience members from media and communications students in Technische Universtät Ilmenau were filmed during a 90 minutes seminar class. There are 6,484 behavior patterns analyzed. A non-representative convenience participant sample (Sverke, 2007) was used.

Students were attending a seminar in the master course in the summer semester 2017. The students were dispersed across the classroom. Every speaker was asked to prepare in advance and give a short presentation (five minutes long) in front of the audience and was part of the audience during the remaining presentations of the other participants.

The final video recordings of the presentations resulted in 180 minutes of coding material: 90 minutes for the speakers' presentations and 90 minutes for the audience recordings. The presentation of the first speaker was not coded due to insufficient light conditions. Out of the 17 audience members, only 14 were eligible for coding, as three persons were seating in the extreme corners of the filming frame and did not fit entirely within.

Verbal communication: total *n* = 47 verbal communications between speakers and audience members were coded.

Paralinguistic messages: n = 156 intonations were coded for speakers and *n* = 3,371 audience reactions to those intonations (including facial expression, gestures, and actions).

Facial expressions: n = 146 speakers' facial expressions, *n* = 149 eye contacts from a speaker, and *n* = 1,790 eye contacts from audience members.

Gestures: n = 143 valid cases were collected for the speaker section and *n* = 1,707 valid cases for the audience members' gestures.

All the sample numbers for each form of interaction separated for speakers and audience members are presented in Table 6.

Table 6

Sample for each form of interaction for speakers and audience members

Forms of interactions	Speakers (*n*)	Audience Members (*n*)
Verbal Communication	31	16
Paralinguistic Messages	156	3,371
Facial Expressions	295	1,790
Gestures	143	1,707
Total	625	6,884

2.3.6. Data collection and data analysis.

Two cameras were used to record the audience members, and three cameras (including web-camera) were used to record the speakers. Two cameras were positioned

in the corners of the classroom, bind the lecture, and were facing the students. Both cameras were oriented to different angles to ensure that as much sitting area is covers as possible. The other two cameras were positioned in passes on the right and left sides of the classroom. The cameras captured speaker from left and right, from a distance, to have a broad view and record all movements of the speaker. Finally, the web-camera was placed at the table in front of the speaker. The camera was capturing the face of the speaker from a closer view. This angle allowed to see facial expressions and mimics of the speaker in proper detail.

The overall total video duration resulted in one hour 28 minutes. Students of Human-Computer Interaction (HCI) summer semester 2017 class agreed to be participants. $N = 13$ speakers made short presentations. $N = 17$ students were participating as audience members.

Content analysis of videos was done using the IBM SPSS Statistics 22 software. Paralinguistic messages, different gestures, and intonations were tested on their dependence on speakers' behavior. Thus, Chi-Square tests were performed to discover if there was a relationship between categorical variables.

For verbal communication, three variables were coded: dialogue initiator, addresser in the dialogue, and type of dialogue. The dialogue initiator variable was measured on a dummy scale with two categories as a speaker and audience member. Addresser in dialogue variable was measured in dummy scale with two categories as one-to-one and one-to-many. Type of the dialogue category was measured in dummy scale with two categories as question and answer. Descriptive tests were performed to see how speakers and audience members interact using verbal communication.

For paralinguistic messages, seven variables were coded: question intonation, incomplete statement intonation, doubtful statement intonation, confident statement intonation, emphasizing statement intonation, volume. All the variables were coded with the dummy scale.

Non-linguistic gestures were measured with four variables: symbolic, deictic, beat; non-manual. Symbolic gestures variable was measured by categorical scale: handwave/salute, thumbs up / down, clenched fist, crossed fingers, clap, hand rubbing, praying hands. The deictic variable was measured with a categorical scale: pointing with the finger to a person. The beat variable was coded with dummy scale: beat and no beat gestures. The non-manual variable was measured with a categorical scale: head movements, shoulder movements, body movements, and body posture.

Non-linguistic facial expressions were coded with six variables: eye contact, neutral, positive, negative, surprised, unclassified. Eye contact was coded with the categorical scale: at the speaker, at an object, at another audience member. While neutral, positive, negative, surprised, unclassified variables were coded with dummy scale.

2.3.7. Coding rules.

Verbal dialogues between a speaker and audience members in the public speaking situations are not an often event occurring. That is why the dialogues were coded following the event sampling. The video materials were observed, and every occurring dialogue was coded. The same procedure was followed for coding intonations.

Gestures and facial expressions are types of behavior that accrue quite often. Such real-life events can result in a large volume of data. Thus, time sampling was used,

and the behaviors event of the speaker and each audience member was coded every 30 seconds. For this, the video file was paused every 30 seconds, and behavior recorded at this moment was coded. After the coding of the recorded material, all the data was transferred into IBM SPSS Statistics 22 for analyzing.

2.3.8. Ethical consideration.

The observation study was conducted according to the American Psychological Association principles of ethics (Smith, 2003). The study was partially conducted with a Master's research module at the Institute of Media and Communication Science at TU Ilmenau in the summer semester 2017. Students participated in codebook development, participated in the study as audience and speakers, and in coding the video materials.

The written consent forms of participation were signed by all participants on the day of the filming seminar. Participants were informed about the confidentiality of the data, that the data will only be used for research, and no personal data will be published anywhere. The contact data was provided for any further questions. No other personal data was inquired. Participation was voluntary without any incentives or rewards.

2.3.9. Reliability test.

There are different groups of coders coding different types of behavior. Inter-rater reliability was measured for each coder group and each category in the codebook. As suggested by Doering et al. (2015), nominal variables were tested via Cohen's Kappa Coefficient, and interval scaled variables using intra-class-correlation. The first Cohen's Kappa was run to determine if there was an agreement between two coders who coded verbal content in dialogues on different variables regarding verbal content and behavior during dialogues. After the first reliability test, several modifications were made in the

codebook as well as more clear coding instructions were formulated. After amendments, Cohen's Kappa Coefficient was used to check inter-coder reliability. Following the Altman (1990) interpretation, there was moderate agreement between two coders for all the categories; all the results are summarized in Table 7.

Table 7

Categories and inter-coder reliability between two coders of the verbal communication in speaker-audience dialogues codebook

Category	Average Cohen's Kappa Value (K)
Dialogue initiator	100.0
Addresser in the dialogue	83.3
Type of dialogue	86.0

Note. N = 47 dialogues in 8 video-recorded presentations.

The second Cohen's Kappa was run to determine if there was an agreement between three coders who coded different variables regarding speakers' paralinguistic messages during presentations. Emphasizing the statement intonation category is not included in the reliability table since it was never coded. The reliability outcomes were sufficient for all categories and are visualized in Table 8.

Table 8

Categories and inter-coder reliability between three coders of the paralinguistic messages in speakers' speech

Type of content	Average Cohen's Kappa Value
Gender	100.0
Nationality	100.0
Confident	41.
Question	93.
Incomplete	71.
Doubtful	930
Volume	100.0
Overall	100.0

Note. *N* = 709 speakers' intonations in 180 minutes of video-recorded presentations. For nominal variables, Cohen's Kappa was calculated whereas for interval scaled variables intra-class correlation was calculated

As part of the coding of paralinguistic messages, the audience members' reactions towards those paralinguistic messages were coded as well. Cohen's Kappa was run to determine if there was an agreement between three coders who coded audience members' behavior when the speaker used intonations. The reliability outcomes were sufficient for all categories and are visualized in Table 9.

Table 9

Categories and inter-coder reliability between three coders of the audience members reactions towards the speaker's paralinguistic messages

Type of content	Average Cohen's Kappa Value
Gender	100.0
Nods	77.0
Headshakes	100.0
Self-grooming	77.
Negative	100.0
Neutral	67
Positive	65
High gaze	75
Law gaze	74
Interaction	100.0

Note. *N* = 1,790 audience members' reactions in 180 minutes of video-recorded presentations. For nominal variables, Cohen's Kappa was calculated whereas for interval scaled variables intra-class correlation was calculated

The independencies between four coders were determined using Cohen's Kappa inter-rater reliability test for speakers' facial expressions and reactions of audience members. For the first pre-test, each coder independently coded the same two speakers and two audience members. For all the six coders combinations, the pairwise Cohen's Kappa test was performed. The average Cohen's Kappa value for each category was calculated. The first pre-test revealed low or not computable values of reliability among

the coders. It can be explained the fact that facial expressions are quite tricky to interpret. Therefore, the additional training of coders was organized, and additional revision of video material sample-coded was done together with all the coders to achieve higher agreement.

Further refinements and alterations of the codebook categories were then examined by the second inter-rater reliability test. For the second test, the coders were better trained as well as the codebooks gave more specific coding instruction. Thus, the second Cohen's Kappa test showed much higher reliability values overall. First, Cohen's Kappa test was performed for the speaker's facial expressions categories. The total results of the second Cohen's Kappa test for all speaker's facial expressions summarized in Table 10. The average Cohen's Kappa value for each category was higher than .40 after the second pre-test, which is sufficient for deductive categories (Altman, 1990; Landis & Koch, 1977).

Table 10

Categories and inter-coder reliability between four coders of the facial expressions in speakers' speech

Type of content	Average Cohen's Kappa Value
Eye contact	95.9
Positive	100.0
Negative	100.0
Neutral	64.9
Surprise	-
Anxiety	100.0
Other/Unclassified	100.0

Note. *N* = 149 speakers' facial expressions in 180 minutes of video-recorded presentations.

Second, Cohen's Kappa test was performed for audience reactions towards speakers' facial expressions coded by the same coders' group. The total results of the second Cohen's Kappa test for all audiences' reactions summarized in Table 11. The average Cohen's Kappa value for each category was higher than .40 after the second pre-test, which is sufficient for deductive categories (Altman, 1990; Landis & Koch, 1977).

Table 11

Categories and inter-coder reliability between four coders of audience members reactions towards the speaker's facial expressions

Type of content	Average Cohen's Kappa Value
Eye contact	55.2
Positive	74.1
Negative	100.0
Neutral	70.7
Surprise	100.0
Anxiety	-
Other/Unclassified	70.5

Note. *N* = 1,790 audience members' reactions in 180 minutes of video-recorded presentations.

The independencies between four coders were determined using Cohen's Kappa inter-rater reliability test for gestures of speakers and audiences' reactions. In a first pre-test, from the eight gestures categories for the speaker, only two categories had a positive result of the agreement. Similar to the speaker's categories, there were several negative values found in different pairs of rates for the audience. Following the results of the Pre-test, four modifications were applied in the codebook.

With the revised and final codebook for speakers' and audience members' gestures, the second Pre-test was conducted, coding the same audience members and speakers. Significant changes in the results were reviled for the second Pre-test, given

the Kappa Value more than .40 (Altman, 1990). First, the speaker's gestures categories are summarized in Table 12.

Table 12

Categories and inter-coder reliability between four coders of the speakers' gestures

Type of content	Average Cohen's Kappa Value
Symbolic gestures	100.0
Deictic gestures	100
Motor gestures	65.0
Head movements	100.0
Shoulder movements	100.0
Body movements	47.4
Body posture	40.5

Note. *N* = 143 speakers' gestures in 180 minutes of video-recorded presentations

Second, audience members' gestures reviled significant changes in the results for the second Pre-test as well, given the Kappa Value more than .40 (Altman, 1990). All results of the second pre-test, both for speaker gestures and audience members' reactions, are summarized in Table 13.

Table 13

Categories and inter-coder reliability between four coders of the audience reactions towards the speaker's gestures

Type of content	Average Cohen's Kappa Value
Attention	60.2
Head Movements	70.2
Body Movements	44.5
Body Posture	67.8
Hand Movements	74.5

Note. *N* = 1,707 audience members' reactions in 180 minutes of video-recorded presentations

Since the reliability was finally adequately established among all coders for all four groups of variables, the final coding could be carried out.

2.4. Results – Speaker-Audience Interaction Patterns

The results part contains findings from observation of the real speaker and audience. The study was conducted as a didactic reason to demonstrate real speaker-audience linguistic and non-linguistic interactive behavior patterns that have been used in a public speaking situation. The codebook was developed to analyze the video materials. Following the social behavior, the framework codebook was divided into four parts, four different communication behavior types. Respectively, the results are presented in four that are summarized. Table 14 summarizes all four communication types as well as speaker-audience interactive patterns with each of them.

Table 14

Summary of Communication Types and Real Speaker-Audience Interactive Patterns in Order as Presented in Result Section

Communication Type	Interactive Patterns
1. Linguistic verbal communication (2.4.1)	Initiation of dialogues
	Addressees in dialogue
	Timing and types of dialogues
2.Paralinguistic message (2.4.2)	Intonations used by a speaker
	Audience's facial expressions and speakers' intonations
	Audiences' gestures and speaker's intonations
	Audiences' eye gaze and speakers' intonations
	Audience's attitudes and speakers' intonations
3.Gestures (2.4.3)	Gestures use by speakers
	Frequency of gestures used by the audience
	Audience's attitude and speaker's gestures
4.Facial Expressions (2.4.4)	Eye contact
	Facial expressions of audience members
	Audience's facial expressions and speaker's facial expressions

2.4.1. Linguistic verbal communication.

Initiation of Dialogues. Speakers initiated dialogues more often than did audience members. Specifically, in 66% of all cases, speakers were the ones who started the dialogue, while audience members triggered 34% of dialogues. Table 15

demonstrates the frequencies and percentages of dialogues initiations from speakers and audience members.

Table 15

Frequencies and Percentages of Dialogues Initiations form Speakers and Audience Members

Dialogue initiator	*N*	%
Speaker	31	66
Audience	16	34
Total	47	100

The speakers usually begin the dialogue with a conclusion of the speech before giving the audience a question (e.g. “Do you have any questions?”) to confirm if the audience would have a question (for the speaker), while the audience usually make questions about the speech.

Addressees in Dialogue. In the majority of dialogues, only one listener was addressed (70 %): Be it an audience member asking the one presenter a question or the presenter answering to the one audience member that had asked a question. However, sometimes the presenter or audience members addressed more than one other person at once (e.g., the whole group). Table 16 summarizes frequencies and percentages of one-to-one and one-to-many addressees in dialogues.

Table 16

Frequencies and Percentages of Dialogue Addressees

Type of addressees	*N*	%
One-to-One	33	70.2
One-to-Many	14	29.8
Total	47	100

Timing and Types of Dialogues. Most commonly, speakers began the dialogues on concluding their speech, asking the audience for any questions (26%); audience members (69%) asked questions regarding the content of the speech (*N* = 47) after the presentation as well. Table 17**Ошибка! Источник ссылки не найден.** shows the types of speech identified in the dialogues (*N* = 47). In general, "questions" were most common at 55%, while "answers" were provided less frequently with 32% of all types of speech. Other types of dialogues were used very rarely (e.g., appeal) so that they were excluded from the final results—questions more often emanated from the audience, and the answers from the speaker.

Table 17

Types of Speech According to the Role of the Speaker

	Types of speech		
Role of the speaker	Questions	Answers	Total
Speaker	12 (46.2%)	14 (93.3%)	26 (100%)
Audience	14 (53.8%)	01 (6.7%)	15 (100%)

2.4.2. Paralinguistic messages.

Paralinguistic Messages Used by Speaker. From all the intonations used by the speaker during the presentations, incomplete statements were most frequently used (n = 80), while questions (n = 26) least frequently used. Table 18 shows the means of how many intonations every speaker used on average.

Table 18

Frequency of Speakers' Intonations (Percentages, Means, and Standard Deviations)

Types of Intonation	%	*M*	*SD*
Incomplete statement	35	6.67	3.20
Confident statement	31	6.08	2.15
Doubtful statement	23	4.5	5.57
Question	11	2.17	1.34

Note. N = 233 (for 12 speakers). Scale from 0 = none of the speakers used intonation to 12 = 12 speakers used intonation.

Audience's Facial Expressions and Speakers' Paralinguistic Messages. First, the audience's reactions to different paralinguistic messages of the speakers were analyzed. It was tested if facial expressions, gestures, eye gazes, and attitudes of the audience are associated with the speakers' intonations. All variables were measured at the categorical level, and the observations are independent. Therefore, Chi-square tests for association were conducted, and standardized residuals were computed for each of the categories except the formal categories. The assumption of the Chi-square test was checked, namely, if all expected frequencies are greater than five (Cochran, 1954). When the Chi-square tests for association between speakers' intonations and audience member's facial expressions were conducted, all expected frequencies were greater than five except for only one single negative facial expression cell. However, the analyzes were run, because it is acceptable to have up to 20% of expected frequencies below five (Field, 2009; Howell, 2010), but the tests then lose statistical power (Field, 2009).

There was a statistically significant difference, $\chi^2(3) = 167.1$, $p < .001$, with a small effect size of Cramér's $V = .22$, between positive facial expressions in the audience and speakers' paralinguistic messages, as well as a statistically significant difference, $\chi^2(3) = 135.6$, $p < .001$, with a small effect size of Cramér's $V = .20$, between audiences' neutral facial expressions and speakers' paralinguistic messages. There was no significant difference, $\chi^2(3) = 4.3$, $p = .226$, between the speakers' paralinguistic messages and the audience's negative facial expressions even though almost half of them were found during incomplete statements (46.7%), probably because the absolute frequencies were rather small. The detailed results of all three Chi-square tests are presented in Table 19.

Table 19

Prevalence of audience member's facial expressions during speakers' intonations use

	Type of Intonation										
	Confident		Question		Incomplete		Doubtful				
Facial Exp.	*n*	%	*n*	%	*n*	%	*n*	%	χ^2	*p*	*V*
Positive	188	40	61	13	123	26	96	21	167.1	<.001	.22
Neutral	903	31	253	9	1045	36	699	24	135.6	<.001	.20
Negative	11	19	2	3	44	73	3	5	4.3	.226	.04

Note. Frequencies within each facial expression conditions add up to 100%.

Exp. = Expressions; The Chi-Square Test of Independence was performed

Audiences' Gestures and Speaker's Paralinguistic Messages. In a second step, audience member's nods, headshakes, self-grooming, and interactions were tested on their association with paralinguistic messages. Again, all variables were measured at the categorical level. Chi-square tests were computed for each of the categories. The test assumptions were fulfilled. The Chi-square test for association between speakers' paralinguistic messages and audience member's actions was statistically significant, $\chi^2(3) = 26.8$, $p < .001$, with a very small effect size of Cramér's $V = 0.09$. However, there was no statistically significant association, $\chi^2(3) = 3.9$, $p = .270$, between audience member's nods and speakers' paralinguistic messages as well as no significant difference $\chi^2(3) = 0.9$, $p = 826$ between audience member's handshakes and speakers' paralinguistic messages, and, finally, no significant difference, $\chi^2(3) = 1.9$, $p = 576$ between audience member's self-grooming and speakers' paralinguistic messages. The detailed results of all four Chi-square tests are presented in Table 20.

Table 20

Prevalence of audience members' gestures while speakers' paralinguistic messages use

	Type of Intonation										
	Confident		Question		Incomplete		Doubtful				
Gestures	*n*	%	*n*	%	*n*	%	*n*	%	χ^2	*p*	*V*
Interaction	11	27	14	35	5	13	10	25	26.8	<.001	.09
Nods	16	27	3	5	22	37	18	31	3.9	.270	.03
Handshake	3	25	2	17	5	41	2	17	0.9	.826	.02
Self-grooming	72	35	19	9	72	35	43	21	1.9	.576	.02

Note. Frequencies within each audience members' actions conditions add up to 100%. The Chi-Square Test of Independence was performed

Audiences' Eye Gaze and Speakers' Paralinguistic Messages. In a third step, the association between audience members' eye gaze and the speakers' paralinguistic messages was tested. Again, all variables were measured at the categorical level. Chi-square tests were computed for each of the categories. The test assumptions were fulfilled. Both Chi-square tests showed no statistically significant association between audience members' gaze and speakers' paralinguistic messages. The detailed results are presented in Table 21.

Table 21

Prevalence of audience member's eye gaze during speakers' paralinguistic messages use

	Type of Intonation										
	Confident		Question		Incomplete		Doubtful				
Eye gaze	*n*	%	*N*	%	*n*	%	*n*	%	χ^2	*P*	*V*
High	736	31	280	12	809	34	555	23	4.9	.177	.04
Low	425	33	120	10	439	34	300	23	7.5	.058	.05

Note. Frequencies within each eye gaze conditions add up to 100%. The Chi-Square Test of Independence was performed

Audience's Attitudes and Speakers' Paralinguistic Messages. Finally, the association between audience members' attitudes and the speakers' paralinguistic messages was tested. All variables were measured at the categorical level. Chi-square tests were conducted for each of the categories. There was no statistically significant association between an audience member's attitudes and speakers' paralinguistic messages. Detailed results are presented in Table 22.

Table 22

Prevalence of audience members' attitude while speakers' paralinguistic messages use

	Type of Intonation										
	Confident		Question		Incomplete		Doubtful				
Audience Attitude	*n*	%	*n*	%	*N*	%	*n*	%	χ^2	*p*	*V*
Attentive	732	31	280	12	802	34	553	23	5.8	.121	.04
Non-attentive	355	32	101	9	391	36	252	23	6.3	.097	.04

Note. Frequencies within each attentive behavior conditions add up to 100%.

The Chi-Square Test of Independence was performed

2.4.3. Gestures.

Gestures Use by Speakers. As it was stated earlier, all speakers' gestures were divided into manual and non-manual gestures. In the manual category, consisting of symbolic, deictic, and motor gestures, there were two completely different occurrences. The first finding shows that deictic and symbolic gestures were rarely used by speakers. It can be concluded that using gestures by the speaker is a natural form of emphasizing and accompanying the speech. Speakers use gestures frequently, and it would thus, considered as normal to use gestures, and can be seen as a positive aspect in terms of interaction with the audience. Detailed frequencies and percentages of using gestures by the speaker are demonstrated in Table 23.

Table 23

Frequencies and Percentages of Gestures used by Speakers

	N	%
Symbolic/ hand rubbing	4	0.9
Symbolic/ showing number with fingers	6	1.3
Deictic/ person	2	0.4
Deictic/ direction or object	4	0.9
Motor/ beat	108	23.1
Motor/ no beat	35	7.5
Shoulders shrug	1	0.2
Body movement	33	7.1
Body posture closed	68	14.6
Body poster open	75	16.1
Head movements	0	0
No gestures	130	27.9
Total	467	100

Head nods, head shakes, and shoulder movements are more of a responding kind of gestures and were therefore not happening when the speakers were giving a speech. For the body movements part, there was about 23% of coded forward, backward, and sideways movement. There are only clear and extreme movements that were coded, so the number can be even higher. Having an open or closed posture was also a very common appearance.

Frequency of Gestures uses by Audience. As part of facial expression analysis, the eye gaze of the audience was analyzed (see Table 24) and used here as identification

of attentiveness or non-attentiveness. It is shown that the majority of the audience members (72%) keeps eye contact with the speaker during the speech of the speaker. Additionally, head movements and body movements of each audience member were coded. There are $N = 41$ cases were coded in the category of head nods. The head nods were used as a form of approval and showed interest in what the speaker was saying.

Table 24

Frequencies and Percentages of Audience's Head Movements

	N	%
Body movement to front/back/side	105	4
Head nods	41	2
Open body posture	1115	47
Closed body posture	592	25
Hand movement – adapters	489	21
Hand movement – raise	12	1
Total	2354	100

Audience members' body movements of all directions, sideways, backward, or forward were very rarely used (6%). Compared to the speaker's body posture, audience members had predominantly open postures (65%). A hand raises from the audience members happened only a few times at the end of the presentations when the questions were asked, adapters happened quite often (28%).

Audience's Attitude and Speaker's Gestures. The audience's reactions to the speaker's gestures were analyzed. It was tested if the attitude of audience members, such

as attentive and non-attentive attitude, is associated with the speakers' gestures. All variables were measured at the categorical level, and the observations are independent. Therefore, Chi-square tests for association were conducted, and standardized residuals were computed for each of the categories except the formal categories. The data was checked for two assumptions. Firstly, observations were independent (Field, 2009). Secondly, the data were tested on how it fits the model by using SPSS Statistics, namely, that all cells should have expected counts greater than five (Cochran, 1954). All expected cell frequencies were greater than five.

There was a statistically significant difference $\chi^2(3) = 153.1$ $p < .001$ and a moderate association Cramér's $V = 0.41$ between the attentive attitude of the audience and speakers' gestures, as well as there was a statistically significant difference $\chi^2(3) = 135.6$, $p < .001$ and a moderate Cramér's $V = 0.45$ association between audiences' non-attentive attitude and speakers' gestures. The results of two Chi-square tests are combined and presented in Table 25.

Table 25

Prevalence of audience members' attitude and speaker' gestures

	Speakers' Gestures Use						
	Using Gestures		Not Using Gestures				
Audience Attitude	*n*	%	*n*	%	χ^2	*p*	*V*
Attentive	1859	79	508	21	5.8	<.001	.41
Non-attentive	116	11	983	89	6.3	<.001	.45

Note. Frequencies within each attentive behavior conditions add up to 100%

The Chi-Square Test of Independence was performed

2.4.4. Facial expressions.

Eye Contact The results in Table 26 demonstrated that the audience tends to look at the speaker. Specifically, 72% keep eye contact with the speaker during the presentation. The audience members, who were not looking were whether looking at the objects such as phone, laptop, table (27.6%) or looking at other audience members when they talk to each other (0.4%).

Table 26

Frequencies and Percentages of Audience Members' Eye Contacts

Eye Contact	*N*	%
At speaker	1,289	72.0
At object	494	27.6
At another audience member	7	4.0
Total	1,790	100.0

Facial Expressions of Audience Members. Table 27 shows that the dominant facial expression performed by audience members was “neutral”. 72 % of the audience members kept their facial expression as neutral during the presentation time. Seven presents of the audience members exposed their positive facial expressions toward the speaker during the presentation time. Negative” and “Surprise” resulted only in 0.4 %. Noticeably, the audience members have also had a variety of unclassified facial expressions.

Table 27

Frequencies and Percentages of Audience Members' Facial Expressions

Facial Expression	*n*	%
Neutral	1293	72.0
Positive	133	7.0
Negative	8	4.0
Surprised	8	4.0
Unclassified	345	19.0
Total	1787	100

Table 28 demonstrates all unclassified facial expressions which were used by audience members. “Facing down” was found as the most common unclassified action (59%). “Covered face” was rated as the second most used unclassified action (35%). Audience members could cover their faces with their hands or with their hair, or the face could be covered by other audience members. The other unclassified facial expressions included such actions as eating, drinking, turning away, yawning, and lip pressed inside. Those unclassified facial expressions resulted in less than three percent.

Table 28

Frequencies and Percentages of Audience Members' Unclassified Facial Expressions

Types of Unclassified Facial Expression	*N*	%
Facing down	205	59.0
Covered face	120	35.0
Eating/drinking	10	3.0
Turning away	8	2.5
Others	2	0.5
Total	345	100

Audience's Facial Expressions and Speaker's Facial Expressions. The association of the audience's facial expressions and different facial expressions of the speakers were analyzed. It was tested if facial expressions of the audience, such as positive and negative, are associated with the speakers' facial expressions. All variables were measured at the categorical level, and the observations are independent. Therefore, Chi-square tests for association were conducted were computed for each of the categories except the formal categories. The test assumptions were fulfilled. All expected frequencies were greater than five. There was a statistically significant difference, $\chi^2(1) = 30.9$, $p < .001$, between positive facial expressions in audience and speakers' positive facial expression with a moderate effect size of Cramér's $V = 0.45$, but there was no statistically significant difference, $\chi^2(1) = 0.2$, $p = .661$, between audiences' facial expressions and speakers' negative facial expression. The results of two Chi-square tests are combined and presented in Table 29.

Table 29

Prevalence of audience members' facial expressions while speakers' facial expressions use

	Audience Members' Facial Expressions						
	Positive		Neutral				
Speaker's Facial Expression	*n*	%	*N*	%	χ^2	*p*	*V*
Positive	18	90	34	26	30.9	<.001	.45
Negative	1	5	4	3	0.2	.661	.04

Note. Frequencies within each speaker's facial expression conditions end up in 100%

The Chi-Square Test of Independence was performed

2.5. Conclusions

The observation study represents the first step in developing a realistic VR public speaking training application. The observation study approached real speaker-audience interactive patterns. The described real speaker-audience interactive patterns summarize how the communication between speaker and audience is performed during the public-speaking situation. A structured, non-participant overt observation study on speaker-audience interactive behavior was conducted. Audience with $N = 8$ speakers (4 female, 4 males), and $N = 14$ audience members (9 female, 5 male), all of whom had given informed consent, was video-recorded with four video cameras. The sample

resulted in a N = 6,484 interactive behavior patterns that were subjected to quantitative content analysis.

The observation study contributed to the research field by closing the research gap on real human speaker-audience interactive behavior patterns.

The study analyzed real-life speaker-audience interactive behavior patterns to implement those behavior patterns later into a virtual public speaking training application later. The following paragraphs describe the summary of the observation study (2.5.1), discussion (2.5.2), as well as limitations and strengths (2.5.3).

2.5.1. Summary of findings.

Verbal Communication. The results obtained in the observation study show how people interact with each other within the dialogues occurring during presentations in a seminar group, a mutually supportive learning community. Speakers are the more active partners in the dialogues. Further, speakers are more likely to focus only on the audience member with whom he/she is interacting at a particular point. Thirdly, during dialogue, speakers and audience members often create questions to interact with each other.

Paralinguistic Messages. During the dialogue, speakers and audience members often create questions to interact with each other. Speakers used many incomplete and doubtful statements during the presentations. The least frequently used intonation was the question, which could be because presenters were asking the audience if they have questions only at the end of the presentation but never during the presentation itself. The use of intonations by the speaker influences audience's behavior. During questions from the speaker, there were less neutral facial expressions and more positive facial expressions.

On the other hand, during doubtful and incomplete statements, there were less positive facial expressions found. This could be a result of the fact that doubtful and incomplete intonations can confuse the listener (McDonald, 2010). During the confident statements, there was no difference found between the facial expressions. Regarding actions, intonations were only found to affect interactions between audience members. No significant influences could be detected for nods, headshakes, and self-grooming.

Gestures. The non-linguistic communications between the speaker and audience were analyzed as well. The study showed that some categories of the speakers' and audience members' gestures, such as beats and hand movements, are more common than the others. The results revealed that using gestures for the speakers, in general, is a natural form of accompanying and emphasizing the speech. The audience members, on the other hand, have an attentive attitude towards the speaker during a speech and do not use obvious movements. Audience members replied with more attention to speakers who used gestures during the speech.

Facial Expressions. Both speakers and audience members were generally polite. They tend to maintain a neutral facial expression while keeping eye contact with each other during most of the presentation time. One action is that the audience members were facing down to look at their mobile phones, notes, laptops, or other objects. Moreover, other extreme facial expressions like "negative" and "surprise" could rarely be observed in audience members' behavior. Audience members replied with positive facial expressions to the speaker's positive facial expression.

2.5.2. Discussion.

This study aimed to find out interactive behavior patterns that should later be implemented into VR public speaking training applications to make it more realistic. The study was based on the social behavior framework and analyzed both the linguistic and non-linguistic behaviors of the audience and speaker and analyzed the association of the speaker's behaviors and audiences' behaviors.

Following the social behavior theory, linguistic behavior is one of two parts of social interactive behavior. The linguistic behavior of the real audience members and real sparkers was analyzed as part of this study. Observation showed how audience members and speakers interact with each other within the dialogues occurring during presentations between speaker and audience. It was found that speakers are the more active partners in the dialogues, they are more likely to *initiate* the dialogues with audience members, often speaker and audience create questions as dialogue *types* to interact with each other, and *addressees* in dialogues were mostly one-to-one (to a speaker or one audience member). These were the first hints for creating an interactive virtual audience. While designing VR training applications, a virtual audience should not initiate dialogues during a presentation and only react to such key sentences as 'Do you have any questions?' at the end of the presentation.

Paralinguistic messages, as part of linguistic behavior, were observed and analyzed as well. Findings supported that most of the time, speakers use question intonations to start a dialogue with the audience. The use of intonations in speakers' presentations influenced audiences' behavior. Question intonation leads to more positive facial expressions, while during incomplete and doubtful statements, there were less

positive facial expressions in the audience. In implementing an interactive virtual audience, speakers' intonations should be taken into account. VR application developers can use these findings later when the intonation recognition tools will be more advanced and manipulate the virtual audience to react on question intonation automatically. At the same time, the volume of the speech should be measured and be adjusted to the level that the audience can hear the speaker. It is supported by literature that speakers' voice loudness correlates with audience engagement (Curtis, Jones, & Campbell, 2015). Although it had never happened during observation, due to the use of high voice volume during presentations, it can be assumed that the low voice volume will affect audience engagement. In the case of contentious use of low volume, the audience might request from the speaker to speak louder.

Second, the non-linguistic behavior of a real audience and the real speaker was observed and analyzed. Gestures are considered as a significant part of any human interaction. To make user interaction within VR applications more realistic, gestures should be taken into account and implemented as a communication part. Overall results, from observing speakers' gestures and their association with audiences' behavior, showing that the use of gestures by the speakers is a natural form of emphasizing and accompanying speech, which creates a more attentive behavior from the audience during a speech. While not using gestures during the speech leads to a non-attentive behavior form the audience. This finding is relevant and useful for developing an interactive virtual audience. The number of gestures used during the speech should be taken into account, and the audience's reactions should be manipulated accordingly.

Finally, facial expressions and eye contact were measured. Both speakers and the audience generally tend to maintain a neutral facial expression while holding eye contact during most of the presentation time. One action is that the audiences were facing down to their mobile phones, laptops, notes, or other objects. Moreover, as expected from the literature, a smile from the speaker was reciprocal and led to a positive facial expression from some audience members. In developing an interactive virtual audience, the virtual audience should be able to express positive facial expressions and keep eye contact with the speaker, as well as react to speakers smile and sometimes smile back.

All the findings for each coded category with recommendations for the further implementations are summarized in Table 30.

Table 30

Summary of findings for each coded category with recommendations for further implementations in VR application

Categories	Findings	Recommendations for the Implementation
Linguistic		
1. Initiator	Speaker	VR audience recognize initiation from a speaker
2. Addresser	One-to-one	One VR audience member participate in dialogue
		Implementation of Q&A session
3. Type	Questions	

Table 30

Summary of findings for each coded category with recommendations for further implementations in VR application

Categories	Findings	Recommendations for the Implementation
Paralinguistic Messages		
1. Question	Often used, leads to positive FE	Implementation of Q&A session with a positive attitude from VR audience members
2. Incomplete	Often used, leads to neutral and positive FE	VR audience members, in general, have neutral and sometimes positive FE as the reaction to the speech
3. Doubtful	Often used, lead to neutral and positive FE	VR audience members, in general, have neutral and sometimes positive FE as the reaction to the speech VR audience members, in general, have neutral and sometimes positive FE as the reaction to the speech
4. Confident	Often used, lead to neutral and positive FE	VR audience members react to low volume
5. Emphasizing	Not used	
6. Volume	Measured with the microphone for technical examination	

Table 30

Summary of findings for each coded category with recommendations for further implementations in VR application

Categories	Findings	Recommendations for the Implementation
Gestures		
1. Symbolic	Rarely uses	
2. Deictic	Rarely used	
3. Beat	Motor most used by the speaker, audience reacts with attentive behavior	Speakers gestures can be recognized and attentive or non-attentive from virtual audience generated accordingly
4. Non-manual	Open body posture mostly used by the audience and speaker	VR audience members should keep open body poster during the speech
Facial Expressions		
1. Eye contact	Towards speaker	Simulate eye contact with the speaker
2. Neutral	Mostly used	VR audience should most of the time have neutral FE
3. Positive	Used randomly and as a reaction to a speaker	VR audience should have positive FE occasionally; few audience members should smile back as a reaction to the speaker's smile
4. Negative	Rarely used	
5. Surprised	Rarely used	
6. Unclassified	No effect	

Note. VR = Virtual Reality; Q&A = Questions and Answers; FE = Facial Expressions

The study managed to reach the goal, the interactive behavior patterns of a real speaker and real audience were successfully observed and analyzed, and the first research question was answered:

RQ1: What patterns of speaker-audience interactions can be identified in real-life?

The next step of implementing realistic speaker-audience interactive patterns in a VR public speaking application prototype can be accomplished next.

2.5.3. Limitations and strengths.

Though the validity and reliability of this study seem strong, it should be noted that these results have some limitations. The first limitations of this study include the fact that the sample consisted of international students from different countries, none of whom had English as a native language. There was the possibility that the listeners did not understand what the speaker was saying and, thus, did not react appropriately to the verbal communication or intonations. However, the study aims to create an application that can be used by internationals for training public speaking skills. Thus, the understanding of international audience behavior is needed.

The second limitation is that the speakers and the audiences were familiar with the setting, the task of the seminar, and with each other, they might have behaved differently in unfamiliar situations with unfamiliar people. Finally, the quality of the video materials was not always good enough. The audience members on the last raw were not always in the field of view. Thus, they could not always be observed clearly.

Due to that, their gestures and facial expressions were not seen entirely and were not possible to code.

Despite these limitations, the study has several strengths. The first study analyzing interactions between real audiences and the real speaker was conducted. The interactive behavior patterns between the real audience and real speaker were observed and summarized. Compared with similar studies that analyzed audiences' behaviors ((e.g., Kang et al., 2013; Tudor & Doering, 2016), this study analyzed both, linguistic and non-linguistic behavior of the audience member and speaker. Interactive speaker-audience behavior patterns were summarized for implementing real behavior into VR public speaking training applications.

3. Implementation Study

Although virtual reality technology has already been successfully used for training public speaking skills, developers aim to design an interactive virtual audience to increase the efficiency of the training. To develop such interactive virtual reality public speaking training applications, the efficient ways of implementing speaker-audience interactive patterns should be studied. The study aimed to find efficient ways and implement speaker-audience interactive patterns defined in the observation study into virtual reality public speaking training applications.

In the following paragraphs, the introduction to the study (3.1), theoretical background on virtual reality public speaking applications design (3.2), methodology of the study (3.3), results (3.4), and conclusions (3.5) presented and discussed.

3.1. Introduction

Virtual reality (VR) is a technology rapidly growing popularity that simulates real or imaginary environments with a high degree of realism and interactivity (Bishop & Fuchs, 1992; Zeltzer, 1992). VR has already been actively and successfully used for training applications since VR training environments allow exposing users to a simulated situation in a safe, highly visual and interactive way (Aldrich, 2004; Ferry et al., 2004). VR training applications, adjusted appropriately, provide a new level of user-system interaction, allowing users to participate actively in virtual situations (Earnshaw et al., 1993).

One context where such applications are used is public speaking training (Bodie, 2010; Furmark, Tillfors, Stattin, Ekselius, & Fredrikson, 2000). However, to be

successful public speaking applications, realistic *speaker-audience interactions* should be implemented (Poeschl et al., 2014), i.e., dynamic sequences of actions between a human speaker and virtual audience members. They modify their subsequent actions in response to their interaction partners (Turner, 1988). At the same time, the ways through which humans gain social skills and how technology can maximize this process are an under-researched domain in human-computer interaction (Slovák & Fitzpatrick, 2015).

It was stated that the training applications involving virtual humans have a disadvantage that the virtual humans and objects in a virtual environment are not 'real', but synthetic and limited in their appearances, behaviors, and their ability of mimics (Ludlow, 1996). Thus, designing a realistic virtual audience implies two main parts: the realistic appearance of the virtual audience (Slater, Pertaub, & Steed, 1999) and realistic interactive behavior of virtual audience (Poeschl et al., 2014).

For creating virtual humans, there two main ways are exit, such as computer-created avatars (e.g., Chollet, Wörtwein, Morency, Shapiro, & Scherer, 2015) or use pictures/videos of real people (e.g., Lee et al., 2002b). Realistic audience behavior is a more complicated task involving the understanding nature of human communication. Most of the existed VR public speaking applications use prepared in advance or scripted scenario (e.g., Kang et al., 2013) or involving operator who is regulating virtual audience's behavior online (Vinayagamoorthy, Steed, & Slater, 2005).

However, there is no independent interactive virtual audience existed at the moment. Modeling virtual characters with high visual appearance realism and realistic behavior is one of the challenging tasks for developers (Goetz, Kiesler, & Powers, 2003; Slater, Sadagic, Usoh, & Schroeder, 2000). Thus, this study aims to implement reactions

of virtual audiences towards the speaker's behavior. With this aim in mind, the second research question was formulated as:

RQ2: How can speaker-audience interaction patterns be implemented into the virtual public speaking application?

Three steps were done to answer the research question and create an interactive virtual audience: (1) implementing new interactive virtual characters, (2) generating interactive behavior manually, (3) using possible autonomous tools (such as speech recognition, gesture recognition, facial expression recognition). The results obtained in the observation study show how people interact with each other during presentations in a seminar group. Those patters were used as speaker-audience interactive behavior and considered and patterns for implementation.

3.2. Theoretical background on VR public speaking training applications design

VR training systems have already been developed since the late 1980s and are a dynamic and upcoming field of research (LaViola, 2017). VR is described as “an approach that uses displays, tracking, and other technologies to immerse the user in a VE [virtual environment]” (LaViola, 2017, p. 8) and represents “synthetic sensory information that leads to perceptions of environments and their contents as if they were not synthetic” (Blascovich et al., 2002, p. 105). VR is frequently used interchangeably with virtual environments (Burdea & Coiffet, 2017). A virtual environment is a spatial and synthetic world, which is generally a 3D, that users can perceive from a view of first-person and that user can control the view in real-time (LaViola, 2017).

VR and virtual environments can address all sensorial preceptors of the human body such as visual sensory, auditory sensory, or gustatory (Blascovich et al., 2002). At

the same time, VR allows users to move within virtual space (Blascovich et al., 2002), be present in the situation, and interact with the virtual objects (Bowman, 1998) or with virtual characters (Gratch et al., 2002). An interactive virtual audience can appeal to all those users' feelings as well.

However, to create a virtual audience that will be able to appeal to all the feelings, appearance, and behavior of such an audience should be comparable to a real audience. Developing such an audience is a complicated task involving an understanding of VR technologies. This paragraph comprises an introduction to the theoretical background technologies and tools used for VR application development. Firstly, the theoretical background of immersive virtual environments, in general, will be described (3.2.1). Later, theoretical background about VR training applications (3.2.2) virtual audience in virtual reality public speaking applications (3.2.3). Finally, the theoretical background of technical implementation (3.2.4) and interactive virtual reality technologies (3.2.5) are summarized.

3.2.1. Immersive virtual environments.

An immersive virtual environment is an artificial environment created with software and hardware that presented to users in a way that looks and feels like a real environment (Baieier, 1993). It is the simulation of the environment (real or imagined) that the user can experience visually in the three dimensions (3D) of width, height, and depth (Isdale, 1998). Immersive virtual environments are also described as technologies that replace the real world by using such synthetic stimuli as 3D visual imagery, imitation of tactile feedback, or spatialized sound.

Immersive virtual environments are designed with an aim. “to let the user experience a computer-generated world as if it were real – producing a sense of presence, or “being there,” in the user’s mind” (Bowman & McMahan, 2007b, p. 36). If the virtual simulation is created and presented with the high-resolution – high quality over all the displays – then the virtual simulation should affect all of the user’s senses (Slater, Usoh, & Steed, 1994). The immersive virtual environment may also provide an interactive experience in real-time with a different kind of feedback like sound, visual, or tactical (Isdale, 1998). It provides a chance for humans to visualize, interact, and manipulate computers and extremely complex data (Isdale, 1998).

Immersive virtual environments are created with VR technology. VR technology can simulate natural stereoscopic viewing. It creates left-eye and right-eye images of a given three-dimensional (3D) scene or object by using a computer (Aukstakalnis & Blatner, 1992). The brain of the VR user integrates the information from two perspectives and creates the perception of 3D space. As a result, VR technology creates the illusion that objects projected on screen have presence and depth. Using VR technology allows users to perceive distance and perceive different objects with spatial relationships between them (Okechukwu & Udoka, 2011).

3.2.2. Virtual reality public speaking training applications.

Fear of public speaking is a social phobia that is prevalent in the general population (Hart et al., 2017; Rothbaum et al., 1995). The leading psychological training method to improve public speaking skills or reduce the fear of public speaking is cognitive-behavioral training, where people are systematically exposed to fear-triggering stimuli (Heimberg & Becker, 2002; Pribyl et al., 2001). In public speaking

training, the main alternatives to real human audiences (presentation in front of a real audience, in-vivo exposition) are imagined audiences (imaginal exposure training) and virtual audiences (VRET).

In vivo type of cognitive-behavioral training is now having a renewal of interest due to the active use of VR technologies for training applications and its associated possibilities (Bouchard et al., 2017). Vivo exposure therapy aims to immerse patients in feared situations in real life (Porter, Porcari, Koch, Fons, & Spates, 2006). VR exposure training can provide more opportunities in comparison to imaginal exposure training (Wiederhold et al., 2002).

Not everybody can realistically imagine the feared confrontation with an audience, while VR technology provides the opportunity to simulate the human audience and the public speaking situation. People with fear can be exposed to virtual fear-triggering stimuli and in VR can role-play in different types of scenarios while training social skills to overcome the fear (Wiederhold & Wiederhold, 1998). Thus, VR training leads to a higher level of elicited fear of public speaking than imaginal exposure training (Wiederhold et al., 2002). To change the structure of fear, the fear must be activated during exposure (Foa & Kozak, 1986). Hence, VR technologies offer a compromising way between in vivo and imagining therapies and are actively used to train public speaking skills and reduce the fear of public speaking.

VR can substitute a real feared stimulus (audience in case of public speaking) with created and controlled ones. VR training applications allows users to face computer-generated simulation similar to a real-life situation to be exposed to fear-inducing stimuli (Parsons & Rizzo, 2008) and provide such new possibilities as 3D

graphics, VR smartphone applications, augmented reality, head-mounted displays and others (Bouchard et al., 2017). These technologies are used extensively with some phobias (e.g. acrophobia, fear of flying), and yet such applications are now extending to more complex disorders (Wiederhold & Bouchard, 2014).

VRET allows us to regulate the amount of exposed fear and control the exposure process in general. At the same time, the VRET, in comparison with in vivo therapy, can be stopped (by therapist or patient) in any time, repeat the experience (Strickland, 1996), and do it with no violation of patient confidentiality (Riva et al., 2001; Strickland, 1996). Additionally, in some cases, like fear of flying, treatment requires less time theme (Wiederhold & Wiederhold, 1998). The VRETs are in the stage of development now and have some disadvantages as well. Sensors pick up the movement of the user's head in VR and send the information to the computer. The computer adjusts and updates the picture on display according to the location of the user in the virtual environment. The adjustment takes some short time, but it causes a lag, and it can lead to a VR sickness (Lewis & Griffin, 1997).

Several studies have already successfully used VR exposure training to reduce fears and phobias, such as fear of flying for example (Rothbaum et al., 1995; Rothbaum, Hodges, Smith, Lee, & Price, 2000), claustrophobia (Botella et al., 1998), arachnophobia (Carlin, Hoffman, & Weghorst, 1997), and others. Fear of public speaking has also been successfully reduced by VR training applications (e.g., Anderson, Zimand, Hodges, & Rothbaum, 2005; Pertaub et al., 2002; Safir, Wallach, & Bar-Zvi, 2012).

3.2.3. Virtual audience in virtual reality public speaking training applications.

Virtual audiences for treating fear of public speaking have already been investigated (e.g.: (Chollet et al., 2015; Slater et al., 2006). Some applications were developed to simulate the real audience and a real classroom environment.

Slater et al. (1999) developed a virtual seminar room, which was matching a real seminar room and populated it with eight virtual agents seating in a semicircle, facing the speaker. Virtual agents could display random autonomous behaviors (e.g.: twitches, blinks, and nods), could display six facial expressions (like yawns and sleeping faces), and simulate the eye contact by looking and move the head towards the speaker (Slater et al., 1999). Sounds, such as clapping, and yawning were included in the program as well. The application was one of the first of its kind and was quite impressive because it could apply and implement the main characteristics of a real audience. However, the main disadvantages of this application were the use of old technology and graphics, which made it look unrealistic (see Figure 4).

Figure 4: Screenshot of Virtual Audience, used by Slater et al, “negative” audience at the beginning of a presentation vs avatar walks out of the virtual room. From “Public Speaking in Virtual Reality: Facing an Audience of Avatars”, by M. Slater et al., 1999, *IEEE Computer Graphics and Applications*, 3-4 p. Copyright 2018 by Mariia Dubiago.

Later improvements of VR public peaking training applications can be divided into two directions: technical improvements (graphic, sound, etc.) and behavior improvements (attentive/nonattentive attitude, interactivity). The technical improvements are depending on innovations and done through such things as graphics improvements, sounds improvements, or improvements of audience appearance. Lee et al. (2002b), to improve the appearance of the audience, used photos of real people from a digital camera for creating a virtual audience member, which made it look more realistic. At the same time, developers tried to improve the realism of the audience as well, by having an operator who selected audience members' actions during the speech. Figure 5 shows the audience created by Lee et al. (2002b), consisted of 13 people and organized as a standard classroom.

Figure 5: Screenshot of the Virtual audience, created by Lee et al. The scene of a virtual reality system. From Virtual "Reality System for Treatment of the Fear of Public Speaking Using Image-Based Rendering and Moving Pictures" by M. Lee et al., 2002, *Cyberpsychology and Behavior*, 3. p.195. Copyright 2018 by Mariia Dubiago.

Improvement of the virtual audience behavior is a more complicated task, which includes the understanding of real audience behavior. Pertaub et al. (2002), from University College London, conducted a study with the virtual audience consisted of eight male avatars. The application was based on the application used by Slater et al. (1999) and described earlier. However, researchers created three different scenarios for the virtual audience: positive, neutral, and negative behavior. The behavior was recreated by using nonverbal communication, such as facial expression, body postures, and movements (see Figure 6). The researchers stated that virtual audiences can display attention just like a real audience through body orientation towards the speaker, eye contact with the speaker, as well as smiles and nodding (Pertaub et al., 2002).

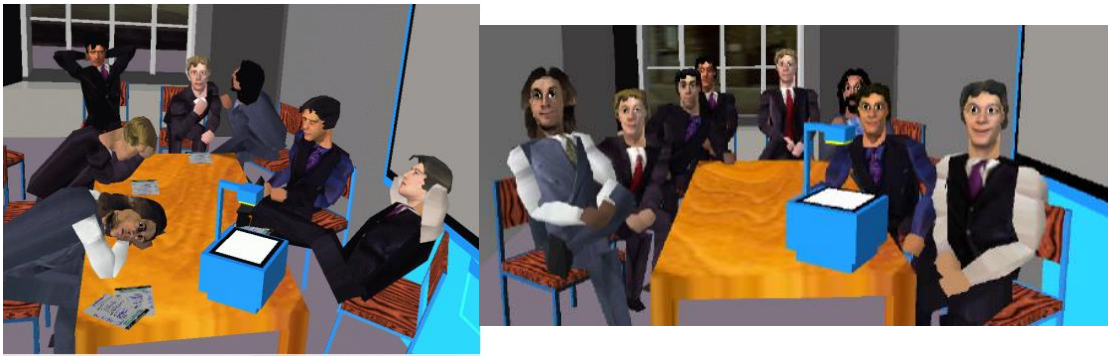

Figure 6: Screenshot of attentive and non-attentive Virtual audiences, used by Pertaub et al. The negative audience on the left vs the positive audience on the right. From "An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience" by D. Pertaub, M. Slater, C. Barker, 2002, *Journal Presence: Teleoperators and Virtual Environments, Volume 11 Issue 1*. Copyright 2018 by Mariia Dubiago.

There are several ways to create virtual audience behavior. Virtual audiences can be preprogrammed in advance (have a script and behave independently from speaker behavior) and/or manipulated by a phobia treatment specialist. Lee and Nass (2005). Besides improving the appearance of the audience, making it more realistic by adding an interaction into the application. It was implemented by using a real-time capture technique. The virtual audience consisted of 12 virtual audience members and one extra character. The extra character is a real-time captured by a PC camera person and inserted into the application. That is why the speaker could interact with a character from a virtual audience, without knowing it was a real person.

Another example of interactivity implementation into the virtual audience is a virtual audience created in Institute for Creative Technologies, Los Angeles, and described in the study by (Chollet, Sratou, Shapiro, Morency, & Scherer) see Figure 7).

Figure 7: Screenshot of virtual audience created in Institute for Creative Technologies, Los Angeles by Chollet et al. study. Screenshot of the running system. From "An Interactive Virtual Audience Platform for Public Speaking Training" by M. Chollet, G. Sratou, A. Shapiro, L. Morency, S. Scherer, 2014, *Proceeding AAMAS '14 Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems, p. 1657. Copyright 2018 by Mariia Dubiago*

This program used different sensors to recognize specific patterns in speakers' behavior. Description of behavior patterns was prepared in advance, and conditions were predefined to manipulate the audience to react to those behavior patterns. Each virtual character's reaction was described for every specific situation. The Behavior Recognizer in this application could extract certain behaviors and run the so-called

Performance Handler, while the Performance Handler accumulated all the behavior descriptions and picked the ones that fitted the description of the Feedback Producer.

In the end, the system was able to display the specific behavior from virtual character, when the particular behavior was observed, and the specific conditions were mat. Thus, the interactivity of the application was much improved; however, the interactivity was limited. Only two speaker's behavior patterns were recognized: gaze (looking or not looking to the audience) and speech, specifically filler words (such as "err", "um" or "uh", that were explained as poor public speaking performance).

Poeschl, Tudor, and Doering (2014) to design and implement realistic audience behavior patterns observed a real audience. Three real lectures with eighteen undergraduates were video recorded and analyzed. The real audience behavior was analyzed to find out interactions only between audience members. The findings from observations were used to improve the realism of the virtual audience's behavior and to make it display nonverbal behavior patterns like a real audience. However, reactions of the audience towards the speaker were not observed and were not applied in this application. It was a simple alpha version prototype with two types of virtual characters included in the scene and with a very limited selection of nonverbal behavior. Thus the application did not look quite realistic (see Figure 8).

Figure 8: Screenshot of virtual audience created in TU Ilmenau for the Public Speaking Anxiety Application (alpha-version) and used by Poeschl and Doering in study. From "Effects of Simulation Fidelity on User Experience in Virtual Fear of Public Speaking Training - an Experimental Study" by S. Poeschl, N. Doering, 2014, *Studies in Health Technology and Informatics*, 199:66-70. Copyright 2018 by Mariia Dubiago.

Two main methods exist to develop the appearance of audience members. One is to use videos or pictures of actual people and embedder those in a virtual environment (e.g., Dubiago, Poeschl, & Doering, 2017; Lee et al., 2002b). In this case, video data should be recorded in advance. Another method is to create computer-created avatars when the audience members look less realistic but have easier construction of behavior database (e.g., Chollet et al.).

There are also two main methods of implementing the behavior of the virtual audience. The majority of the current virtual audience public speaking applications are represented by 3D models that are animated by the predefined script (Slater et al., 2006, e.g.,) (Slater et al., 2006, e.g.,). Animations or videos are usually prepared in advance and manually manipulated by operators in real-time. Such a creating process of the virtual audience required considerable time and effort. That is why the behavior of the virtual audience is often prepared in advance, and a certain scenario is scripted for each virtual audience member. Pre-scripted animations involve long-time preparation and

considerable effort, and that why they are generally relatively short and cause repeating loops in virtual audience behavior (Kang et al., 2013). Repeating loops often reduce the realism, treatment efficacy, and lead to lower effect (Vinayagamoorthy et al., 2005).

3.2.4. Audience's responsive behavior.

Different audience's interactive behavior can be combined in a group and described as the audience's responsive behavior towards the speaker. The responsive behavior of the audience is recognized by non-linguistic behaviors that indicate interest or lack of interest in the speech. Three main kinds of audience responsive behaviors were analyzed for this study: attentive, non-attentive, and neutral.

Attentiveness is defined as a combination of linguistic and non-linguistic communication signs that people use to express that they are understanding and noticing someone's message. In speaker-audience interaction, there are several markers of perceived displayed attention, such as nodding, smiling, direct eye gaze towards the speaker. Body orientation, leaning towards a speaker (Norton & Pettegrew, 2009), and mimicry (Gueguen, Jacob, & Martin, 2009).

For audiences, attention is related to direct eye gaze towards a speaker, little body movements, and note taking (Wilson & Korn, 2007). Mehrabian (1971) stated that "people are drawn toward persons and things they like, evaluate highly, and prefer; and they avoid or move away from things they dislike, evaluate negatively, or do not prefer". The attentiveness of the audience can be recognized through such non-linguistic actions as increased eye gaze, physical proximity, positive facial expressions, body lean towards the speaker (Myers & Ferry, 2001). Attentive attitude of the speaker is studied more than the audience's attitude (Baringer & McCroskey, 2000). However, it has been

studied that non-linguistic attentive behavior occurs in audiences as well, and the attentive reaction is happening the same way as from speaker through eye contact and leaning forward towards the speaker (Kang et al., 2013).

For this research, such non-linguistic attentive patterns were chosen as eye gaze/eye contact, front-oriented body position, nodding (Richmond, 2003), taking notes, and smiles (Wilson & Korn, 2007).

Non-attentive behavior characterized as avoidant behavior such as not looking to the speaker, but somewhere else or being busy with activities that are not related to the speech and speaker (Mann & Robinson, 2009). On the one hand, attention from the speaker can be disturbed by other events. In this case, non-attentive behavior occurs while audience members' attention from original gaze orientation (speaker) starts to focus on another stimulus than the speaker (Remington, Johnston, & Yantis, 1992). As a result, audience member performs speaker-avoiding behaviors such as body posture directed away from the speaker or eye gaze away from the speaker (Mann & Robinson, 2009).

On the other hand, non-attentive behavior can occur due to the broadness of the audience members. Vogel-Walcutt et al. (2012) stated that "boredom occurs when an individual experiences both the (objective) neurological state of low arousal and the (subjective) psychological state of dissatisfaction, frustration, or disinterest in response to the low arousal" (p.102). Audience members use different activities to deal with broadness during the speech, and start using electronic devices, talk to each other, texting on phones or computers, daydreaming (Mann & Robinson, 2009).

Finally, besides attentive and non-attentive behaviors, there is another type of behavior that has neither positive nor negative connotation. It is called neutral behavior. Such behavior does not have any connotation to the speech or speaker. Such activities as playing with hairs, drinking water, scratching, or eating can be attributed to neutral behavior (Beattie & Shovelton, 2016).

3.2.5. Technical Implementations.

In recent years several improvements have been made in virtual audience implementation. At the same time, the success of a virtual audience in training pubic speaking skills and reducing public speaking fear has been positively proven. Nevertheless, developing an interactive virtual audience and creating a successful training application is a complex endeavor that requires competence and expertise in different areas like computer animation, speech synthesis, or multimodal perception. Although realistic virtual characters still cannot compete in realism level with actual humans (Kwon, Powell, & Chalmers, 2013), some findings suggest that virtual audience can reduce the fear that related to the perception of being judges (Lucas, Gratch, King, & Morency, 2014), or increase willingness to train and lower emotional barriers (Hart et al., 2017).

Still, the quality of the virtual audience is highly dependent on the technology and technical equipment that is used for creating a virtual environment. Four leading technologies are needed for creating VR applications: (1) graphics rendering, (2) displays, that immerse users in the virtual environment and block them from the real world, (3) the database construction and maintenance system, to create realistic and detailed models, (4) tracking system that allows continuously identify the position of the

user in virtual environment and operation of users head and other limbs (Burdea & Coiffet, 2017; Durlach & Mavor, 1995).

VR technology consists of two components: software and hardware (Baieier, 1993). The hardware consists of five components: input devices, tracking system, computer workstation, process acceleration cards, and sensory displays (Okechukwu & Udoka, 2011). (1) Input devices are allowing users to interact with the virtual environment. Input devices send signals about users' actions to the system and provide reactions back to users in real-time. Examples of such an input device can be a joystick, keyboard, or voice recognition. (2) To track the orientation and positions of the user in a virtual environment, the tracking system is needed. The tracking system refers to as position sensors (Dani & Rajit, 1998). (3)Computer workstation refers to a high-end computer that is specially designed for technical or scientific applications (Dani & Rajit, 1998; Okechukwu & Udoka, 2011). (4) The process acceleration card helps to update the display according to updated sensory information (e.g., 3D sound cards, 3D graphic cards) (Okechukwu & Udoka, 2011). (5) Finally, sensory displays can display the simulated virtual environment to users. Typically, there are two main devices used to display immersive virtual environments: head-mounted displays and room-sized environments (Blascovich et al., 2002). Such devices provide high technical fidelity, immerse users to situations similar to real-life situations, and provide a feeling of being in the simulated virtual world (Slater, 2009).

There are several displays setup, which is used to immerse people in a virtual environment, those are: head-mounted displays, CAVE (Cave Automatic Virtual Environment) surround projectors, panoramic projectors, workbench projectors, and

desktop displays (Brooks, 1999). Head-mounted displays and CAVE are two display setups that used most often nowadays.

Head-mounted displays include such devices as Oculus Rift and describe as helmet-like devices that can show 3D images (Shibata, 2002). Head-mounted displays use glasses and headset to immerse the user into the virtual environment and completely isolate users from the real world.

CAVE was introduced for the first time in 1992. However, it was not used in exposure treatments until 1997 (Strickland, Hodges, North, & Weghorst, 1997). Room-sized environments include CAVE and defined as "a multiperson, room-sized, high-resolution, 3D video, and audio environment" Waly and Thabet (2002, p. 151). CAVEs usually build from three or six lateral, ceiling, and floor mobile wall (Waly & Thabet, 2002). 3D images are being projected on walls and can be seen with special glasses (like shutter or polarized) (Waly & Thabet, 2002).

Additionally, usually, users wear a position tracking sensor, which recognizes the location of the user in a CAVE and changes 3D image according to the user's position. CAVEs allow several people to use virtual applications at the same time and same physical space. However, the images will only be adapting to the user, wearing tracking sensors. As an example of the CAVE, the CAVE used in this study demonstrated in Figure 9. This CAVE has three frontal projection walls (2,8 x 2,1 m each), with a total length of 8,4 m and with a total height of 3,24 m. The gap between the walls is less than 1mm (FASP, 2014). The whole setup is located on the campus of Technische Universität Ilmenau, Germany.

Figure 9: Three position adjustments possibilities of CAVE screens (90°/135°/180°). Retrieved from https://www.tu-ilmenau.de/komwi/media-box/ab-medienproduktion-sose-2007/faszination-virtual-reality/. Copyright by Mariia Dubiago

Another advantage of CAVE is that virtual environments and real environment can be merged inside of it. However, there are also few disadvantages of using CAVE as VR display setup that can be highlighted such as: due to the big screen size the brightness level is limited, there is a big space that is requirements for the projection, there are a reduced contrast and color saturation, and last but not list, edge effects between the screens can intrude into the displayed scenes (Brooks, 1999).

Applications that use a CAVE setup include the multi-process program, including Tracking Process; Display Process; and Application Computation Process (Usui & Omura, 2007). The tracking system aims to detect the user's positions and directions by using sensors (it can be a wand or glasses with the sensors on top of it) in real-time while the VR application is running. The computer can retrieve the user's positions from the tracking process. Finally, the image is being displayed on the screen of the CAVE according to the user's position.

3.2.6. Interactive virtual reality technologies.

Picard (1999) started that "affective computing and Human-Computer Interaction research target four broad areas: a) reducing user frustration, b) comfortable

communication of user emotion, c) infrastructure and applications to process effective information and d) building tools to support the development of socio-emotional skills" (p.829). Designing a successful VR training application involves understanding of

To design the successful VR training application with a virtual audience being able to respond to the speaker, the application has to recognize users' emotions and behavior. To do that, different recognition tools can be applied (Hudlicka, 2003).

3.2.6.1.Speech Recognition for Verbal Communication

Verbal communication is general and the most natural way of communicating. Verbal communication is also considered the most efficient form of exchanging information among humans. (Shanbhogue, Kulkarni, Suprith, Tejas, & Nagarathna, 2016b). The interaction between humans is mostly done by verbal communication, therefore, it is normal for humans to expect some speech interfaces with computers (Shrishrimal, Deshmukh, & Waghmare, 2012). Thus, researchers are trying to develop a technology with natural language speech recognition for HCI.

Speech recognition is defined as "the process of converting a speech signal to a sequence of words by means Algorithm implemented as a computer program" (Gaikwad, Gawali, & Yannawar, 2010). Speech recognition is also called Automatic speech recognition (ASR) and aims to transfer spoken words into corresponding meaningful text (Shanbhogue, Kulkarni, Suprith, Tejas, & Nagarathna, 2016a).

Speech recognition is a tool developed for human-computer interaction. Humans can differentiate speech very easily. They can make sense of acoustic, linguistic, and context. However, research shows that the relation between corresponding words and the physical speech signal is a very complicated and complex process that is hard to

understand, describe, and especially to implement (Gulzar, Singh, Rajoriya, & Farooq, 2014).

One of the first ASR was developed in 1952 by Davis et al. (1952) in Bell Telephone Laboratories. It could recognize for a single speaker isolated digit from zero to nine. The first ASR systems were based on comparing the speaker's input with acoustic patterns that were predefined (Levis & Suvorov, 2013). This technic is well used for recognizing phonetical distinct items and can operate in with small vocabularies. It cannot be used effectively for larger vocabulary recognition. Additionally, such a system is unable to match input speech with predefined patterns of different lengths (Rabiner & Juang, 1993).

Consequently, this system could not be used for fluent speech recognition. In the past 10-20 years, however, the field of ASR has reached several significant advances based on advances in algorithms, signal processing, computational architecture, and hardware. Nowadays, there is a variety of different speech recognitions that exist. All of the developed ASR systems can be divided into four main types differentiated by kinds of utterances they have the aim to recognize: isolated word (as described earlier very first ASR), connected word, continuous speech, and spontaneous speech (Das, 2012).

Although the public speeches are mores of the time prepared in advance, for the speech recognition tool, they can be considered and spontaneous speeches, since speakers use natural language that includes all speech qualities mentioned earlier. At the moment, spontaneous speech recognition accuracy is rather low (Furui, 2003).

One of the main reasons for low accuracy of spontaneous speech recognition is that all used until now, acoustic and language models have been built based on written

language or speech read from of text. Nevertheless, written and spontaneous speech languages are very different. In spontaneous speech speaking rates are much faster, speaking rates can change even within one sentence, there are a lot of prosodic features, and spectral distribution of phonemes is reduced compared to read speech (Furui, 2005).

Developed in future recognition of spontaneous speech will mean a shift of paradigm from simple speech recognition and transcribing all the words to a more complicated process of actually understanding the information and message of the speaker (Juang, 1998). There are many projects currently working on an increase in recognition coverage in spontaneous speech (e.g., Furui, 2005) (Furui, 2005). However, yet such a tool that could be used for public speaking training application is not available.

3.2.6.2.Emotions recognition in speech for intonations

It is a well-known fact, that the speaker during linguistic communication not just produces a raw textual information message, but transmit a large set of information that can enhance and modulate the meaning of verbal content (Ananthakrishnan & Narayanan, 2009). In a linguistic form of communication, this information is being conveyed by intonations (paralinguistic communication). Paralinguistic features in communication are related to different ways of sentence utterance according to attitudes, intentions, and speaking styles. Such paralinguistic features are related to the emotional state of the speaker (Kleckova & Mahdian, 2004).

Intonations in the context of VR public speaking training applications have already been investigated before, but with a different focus. Often, in paralinguistic speech recognition research, intonation recognition refers to emotion recognition in

speech. In general, research about intonations and emotions recognition in speech can be divided into two aspects: the first, where a machine is designed to produce emotional sounds artificially, the second aspect deals with a machine that recognizes emotional states of a user (Lugovic, Dunder, & Horvat, 2016). To be successful and realistic, machines need to learn about human emotions in a speech by creating a database of real speeches (Lugovic et al., 2016).

Interactions, as a characteristic of single implemented conversational agents, have been studied as the first relevant aspect of successful HCI (e.g., Pelachaud, 2005). Qu (2014) analyzed participants' reactions towards virtual avatars, which were expressing different emotions. The study focused on the emotions delivered by different intonations. "The intonation of the voice was therefore also considered as an important aspect of virtual human's emotional expression of this study" (Qu, 2014). However, intonations in the study were limited to positive or negative (Qu, 2014).

However, most of the research on the recognition of speech focus on emotional recognition in speech. Recognizing emotions in a new and ongoing area of investigations. Manly such research is done in domains of call canter environments. Dellaert et al. (1996) made one of the first research of that type and could identify 60-65% patterns among happiness, anger, and fear in the general domain of HCI. Another research tried to distinguish between negative and positive emotions in a call canter environment and could achieve at maximum the accuracy rate of 75% (Lee, Narayanan, & Pieraccini, 2002a).

Paeschke and Sendlmeier (2000) tried to apply emotion recognizer in real-time by using natural networks, which he adopted for call canter environment conditions and,

as a result, reported a 7% accuracy rate in differentiating agitation and calm emotions. Yu et al. (2001), in their research, tried to detect one emotion among the others and could achieve a 73% accuracy rate for anger, happiness, sadness, and distinct natural emotions. Much other research also focused on recognizing emotions in speech based on real scenarios. For example, Yamamoto et al. (2017) analyzed baby voice, Sharma et al. (2002) used movies as voice databases.

Nevertheless, most of the studies on intonation recognition were focused on call canter environment or environment not similar to public speaking one, and thus, were limited in emotion recognition in speech. At the same time, Chollet et al. (2015) asked participants to experience and evaluate one of the three different virtual audiences: non-interactive virtual audience, the audience that gave direct visual feedback, and the audience that was giving direct nonverbal feedback. The comparison between three conditions was made by self-assessment questionnaires, expert assessments, the measurement of eye-contact, and measurement of pause fillers. The results rated the interactive condition of the virtual audience as the best. Participants in interactive conditions had higher involvement, and public speaking skills were improved by using fewer pause fillers and by using the right intonations (Chollet et al., 2015). The findings of this study prove the importance of the intonation used by speakers.

There are some commercial projects that provide a more extensive range of recognition of speech tools, such as Beyond Verbal, Vokaturi, EmoVoice, and Good Vibrations (Garcia-Garcia, Penichet, & Lozano, 2017). Beyond Verbal recognition, the tool is available by payment that was not affordable for this project to use it (Beyond Verbal, 2017). Vokaturi tool reflects the state of research in emotion recognition in

speech and able to recognize six primary emotions (happiness, neutrality, sadness, anger, and fear) but no intonations. Vokaturi cannot be used in real-time, and some comparison research stated that it gives rather poor results (Garcia-Garcia et al., 2017).

EmoVoice developed for recognition in real-life, but it is a complicated tool that implies developing the model and training (Universität Augsburg University, 2017). It can take too much time to create their databases and configure all programs, especially for a small project (Garcia-Garcia et al., 2017).

Finally, Good Vibrations do not offer any trial version and available only by payment (Good Vibrations, 2017).

It can be summarized that intonations can be used to express the verbal emotions of virtual characters (Qu, 2014). Also, virtual interactive audiences can help improve a presenters' usage of intonations (Chollet et al., 2015). Nevertheless, the implementation of virtual audience reaction towards the speaker's utilization of intonations still lacks research. Human intonations in speech are quite a complicated thing to determine and to create a database of. The meaning of different emotions and their connection with verbal content is still a controversial topic.

Robust recognition of emotions in speech, at the same time, requires a deep understanding of verbal and verbal content aspects of speech in combination (Lee & Narayanan, 2005). Researchers claimed that there is an exact correspondence between positive and negative emotion and a word (Lee & Narayanan, 2005). There are several successful emotion recognition in speech tools that exist; however, in public speaking situations speaker expected to talk professionally without emotions but to use rather intonations. Thus, the intonation recognition topic is an ongoing discussion that

involves many aspects of human communication and interaction, and the solution is not yet developed.

3.2.6.3.Gestures recognition

The gestures, during daily life, are naturally used during communication. However, human-computer interaction systems based on gestures and gesture recognition have various application scenarios. For example, such systems are used in games which are intending to provide experience free from mouse and keyboard (e.g., in (Kratz, Smith, & Lee, 2007); to control robots in some environment (Ma, Xu, & Wang, 2013); or for people who use gesture languages to provide translation (Neiva & Zanchettin, 2016).

Gesture recognition tools refer to the mathematical way of interpreting human motions by a computing device. Gesture recognition recently has become a hot topic, and researchers try to apply gesture recognition for different types of applications. VR interactive applications use gestures recognition to enable realistic manipulations of virtual objects by using hands or create realistic interactions with virtual humans (Gandy, Starner, Auxier, & Ashbrook, 2000; Sharma et al., 1996).

Nowadays, two main technical methods are used to perform gesture recognition (Khan, 2012). The first one is based on wearable electromagnetic devices, such as Nintendo Wiimote (Schlömer, Poppinga, Henze, & Boll, 2008), data glove (Yin et al., 2018), etc. The second one utilizes computer vision and involved image processing (Samantaray, Nayak, & Mishra, 2013). Such technologies are quite expensive and relatively limited, but at the same time, they perform very well (Xu, 2017). For computer vision methods such as advanced technologies as Microsoft Kinect sensors are

used (Biswas & Basu, 2011). This technology is considered more comfortable to use, natural, and less costly compare to technologies mentioned before (Wysoski, Lamar, Kuroyanagi, & Iwata, 2002).

Microsoft Kinect was launched in 2010 and was used as one of the most popular game controllers. It allows users to interact with a computer naturally by computer of a console with gestures and/or voice commands. Microsoft Kinect can interpret information from the 3D scene by using a projected infrared structured light. Light Coding system (3D scanner) makes use of a variant of image-based 3D reconstruction. System sensors are a horizontal bar that is connected to a small base. Sensors have a motorized pivot and designed in a way to be placed lengthwise below or above the video display. The example of Microsoft Kinect is demonstrated in Figure 10.

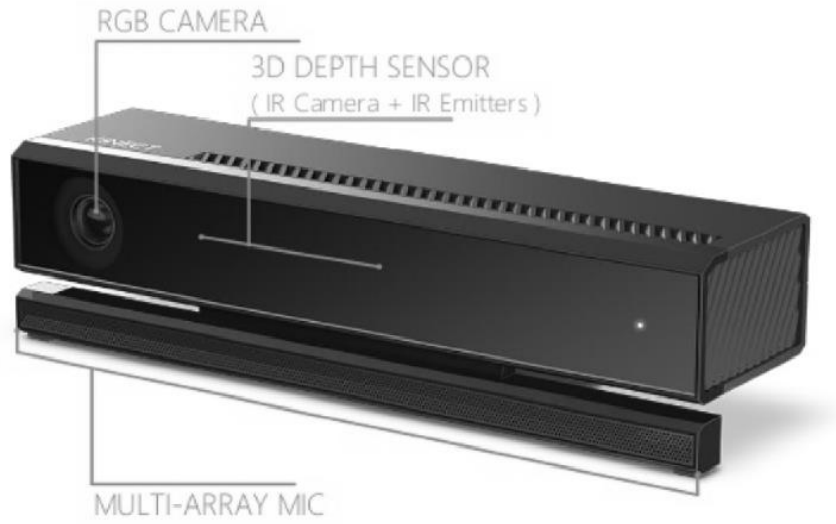

Figure 10: Microsoft Kinect Sensor V2 device. From " Microsoft Kinect in Gesture Recognition:A Short Review" by Soumi, Subhadip, and Mita (2015). In *International Science Press, p. 2072. Copyright 2019 by Mariia Dubiago*

Much work has been reported for gesture recognition. Some research reviews explain the high importance of gesture recognition tools and their use in everyday life (e.g., Khan & Ibraheem, 2012), and especially its importance for HCI, games, robot control and observation of it by using different tools and algorithms (Moesl & Granum, 2001). Most of the works about gesture recognition are based on the use of RGB (color) cameras. Often, gesture recognition attempt to track certain body part like head (Madabhushi & Aggarwal, 2000), to track the joints (Ali & Aggarwal, 2001; Uddin, Thang, & Kim, 2016), to identify particular activities (Rui & Anandam).

3.2.6.1.Facial expression recognition

Just as in real life, nonverbal communication plays a great role in HCI and can be expressed through recognizing and simulating facial expressions and gestures. Such nonverbal aspects of communication are vital aspects of successful public speaking and should be taken into account while designing interactive virtual audiences so that virtual audiences can provide indirect feedback by signaling nonverbally during presentations (MacIntyre, Thivierge, & MacDonald, 1997).

The ability of people to recognize faces is remarkable. Interactive recognition systems to be successful nowadays require the ability to classify a wide range of facial expressions that can be visually distinguished. Humans' faces are complex visual stimuli that are not easy to describe by simple shapes or patterns. Just like a speech, a face can reflect a person's emotions. All the face parts, including nose, lips, eyebrows, mouth, and muscles on the face, are revealing emotions that person is feeling.

The technologies used for facial emotions recognition work analogically to technologies used with speech, and the research can also be globally divided into two

directions: implementation of facial expressions for virtual avatars and recognition of users' facial expressions. On the one hand, the analysis of facial expressions is essential for VR applicators to create the mimicry of avatars. Notably, in VR public speaking training applications, the facial expressions from audience members are recognized as a nonverbal reaction to the speaker and can contribute to stimulating the trained speakers and analyzing facial expressions, thus, necessary to get advanced VR public speaking training applications.

On the other hand, the detection or user's facial expression is needed. In the case of VR public speaking training, speakers' facial expressions should be identified to understand his/her emotions and display a specific reaction from the virtual audience towards it. However, recognizing facial expressions is a complicated thing to do, and it consists of several parts. Firstly, the face itself should be detected. Secondly, the crucial points that explain the emotion should be identified, finally, processing the information and identifying the specific emotion that is being detected.

Facial expressions recognitions are aimed to detect human emotions based on facial expression. Facial expression recognitions generally follow pattern recognition technic and consist of three main steps, such as face detection, feature facial expression itself, and classification of the expression (Chavan & Kulkarni, 2013). Humans' face is the most exposed part of the body, which implies computers video system (such as different video cameras) to analyzes face image to recognize emotions.

Two main factors can affect the quality of any emotion recognition tool using cameras those are: light and change of the head positions (Li, Mian, Liu, & Krishna, 2013). Generally, recognition of changes in the facial expression is based whether on

minor changes such as wrinkles and bulges (Tian, Kanade, & Cohn, 2001) or more significant changes such as nose, eyes, eyebrows, etc. moves (Fasel & Luettin, 2003).

There are two Evaluations, such as the Face Recognition Vendor Test (FRVT) (Grother, Micheals, & Phillips, 2003) make it clear that the current state of the art in face recognition is not yet sufficient for the more demanding applications. Bowyer et al. (2004) surveyed multi-modal face recognition and state of the art in 3D. They summarized the recent research trends as: “the variety and sophistication of algorithmic approaches explored are expanding” (Bowyer, Chang, & Flynn, 2006). There are some challenges in developing face recognition tools, such as improvement of recognition accuracy, the efficiency of algorithms, and the higher robustness of facial expressions (Turk, 1991).

Just as for speech recognition, there are several commercial projects that are developing tools for facial recognition. Several facial recognition tools have been successfully launched and can be used. Those programs are Emotion API (Microsoft Cognitive Services), Kairos, nViso, and Affectiva.

Emotion API (by Microsoft Cognitive Services) can identify face and emotion expression on video or photo (Microsoft, 2017). The system can recognize six basic emotions based on Ekman’s (1977) classification and contempt and lack of emotion (natural facial expression) additionally. The tool is only recognizing facial expressions on uploaded pictures or videos. No real-time recognition included. The system is paid. The demo version includes recognition of face emotion on a picture, which is not relevant for this research.

Kairos tool includes such services as face identification and verification, face detection, facial emotion, gender, age, ethnicity, sentiment, multi-face detection, attention measurements, and face grouping (Kairos, 2017). Again, the tool is working with uploaded pictures and videos, where the video is being sliced into 0.033 seconds-segments, where each segment is being analyzed to find faces and facial expressions on each face. Kairos is working with uploaded materials, and the results are not immediately available. Thus, the tool cannot recognize real-time behavior. Several licenses are available for payment. The free version is only available for personal purposes. Also, it was reported that the result of this recognition tool might be wrong (Garcia-Garcia et al., 2017).

NViso, as stated on the official website, "provides the most scalable, robust, and accurate artificial intelligence solutions to measure instantaneous emotional reactions of consumers in online and retail environments" (nViso, 2016) and includes detection of six primary emotions based on Ekman (1977) classification and natural facial expression additionally. However, the price information should be requested personally, and it does not have any demo or trial version.

Finally, Affectiva recognition tool can detect seven emotions, gender, age, and ethnicity (Affectiva, 2016). The system recognizes 20 zones of the face and them by pixels. Affectiva then classifies detected facial expression following Paul Ekman's Facial Action Coding System (1978). The huge advantage of the system is that it is available for free for academic purposes. Thus, this system was tested for this study.

It can be summarized that facial expressions can be used to express verbal emotions of virtual characters, and users' facial expressions can be used to create an

interactive virtual audience. The implementation of the virtual audience reaction towards the speaker's utilization of facial expressions still lacks research. Robust recognition of facial expressions on humans' faces requires a deep understanding of mimics, and most of the existed programs use Paul Ekman's Facial Action Coding System Emotion detection through facial expressions is a growing research field.

In the last two years, the facial expressions recognition topic became very popular, and at the moment, there are hundreds of companies working specifically on it—researchers investing in resources to build effective applications. Although emotion detection through facial expressions still has many aspects of improving, such toll as Affectiva can be already effectively used by researchers for academic purposes.

3.3. Methodology

The following paragraphs describe the research question (3.3.1) and the design of the implementation study (3.3.2). The design includes such subparagraphs as a description of VR public speaking training application at TU Ilmenau (3.3.2.1), description of the hardware of the application (3.3.2.2), and description of the software (3.3.2.3).

3.3.1. Research question.

The study aimed to implement interactive behavior patterns into the virtual audience. The interactive speaker-audience behavior patterns were observed and analyzed in an observational study. However, the ways of applying this knowledge into application development are not researched. Most of the existed public speaking training applications do not provide interactive autonomic behavior; at the same time, new developments in technologies offer a variety of different recognition tools. Some of the

recognition tools might be relevant to the interactive virtual audience development. Those tools should be researched and tested. Thus, the second research question of this dissertation thesis was formulated as:

RQ2: How can speaker-audience interaction patterns be implemented into the virtual public speaking application?

The public speaking training application prototype developed by Dr. Drumm at TU Ilmenau in Kompetenzzentrum Virtual Reality was used for further interactivity implementation. To answer the research question, three steps have been accomplished: (1) new interactive virtual characters were implemented into the existed public speaking prototype, (2) control panel for regulating and adjusting interactive behavior manually was developed, (3) possible recognition tools (such as speech recognition, gesture recognition, facial expression recognition) for autonomous interactive behavior of the virtual audience were applied. The study has an exploratory approach and belongs to descriptive research (Vaus, 2001). The investigation was done without having previously explicit expectations (Engel & Schutt, 2014).

3.3.2. Design.

Known from a theoretical background, an immersive virtual environment is an artificial environment created with software and hardware (Baieier, 1993). This paragraph will describe the application used as the basis for interactivity

implementation, hardware, and software used in this study to create an interactive VR public speaking training application.

3.3.2.1.VR Public Speaking Training Application at TU Ilmenau

The VR public speaking training prototype was developed earlier at TU Ilmenau in the Kompetenzzentrum Virtual Reality by Dr. Drumm, and existed prototype was used for further development. Developed by Dr. Drumm virtual audience consisted of $N = 12$ people (6 male, 6 female). The Arbitrarily Shape Video was used to represent the virtual agents within the audience, as it works in a standardized scene description mixed with 3D-objects in one common profile (Drumm, 2002; Kühhir, Drumm, Reiter, & Rittermann, 2002; Pereira & Ebrahimi, 2002).

The recording of Arbitrarily Shaped Video virtual characters is fast and creates non-synthetic agents. To obtain natural-looking agents in the present instance, each member of the virtual audience was created by video recordings of a real person. Each virtual character had up to $N = 10$ behavior animations in the video database.

All behavior animations were assigned into four groups: postures, actions, interactions, and reactions. The postures included different kinds of positions, such as leaning back, leaning front, or leaning right. The actions were defined as any kind of activities of virtual audience members', excluding conversational actions, and included writing notes, drinking water, scratching head, and others. The interactions were defined as conversations between the virtual audience members. The reactions were defined as interactions between speaker and audience members, such as questions from the audience. Table 31 presents the four animation categories with examples.

Table 31

Animation categories with examples

Category	Animation Examples
Postures	Lean back/forward, lean right/left, cross arms
Actions	Check mobile phone, type on a computer, write notes in a notebook, clean nose, drink water
Reactions	Ask the speaker to talk lauder, ask the speaker to repeat the sentence
Interactions	Audience members talk to each other about plans for the evening, asking the speaker questions related to the topic

All virtual characters can individually and randomly change their postures and display different actions, interactions, and reactions. The audience behavior of the prototype was not prepared in advance and was not regulated by an operator but was generated spontaneously. However, the virtual audience's behavior did not depend on the speaker's behavior. Further implementations of the described virtual audience were done in collaboration with Dr. Drumm at TU Ilmenau in the Kompetenzzentrum Virtual Reality. The technical side of the implementation study was developed by Dr. Drumm, while the conceptual and design part is explained within the results paragraph of this chapter.

3.3.2.2.Hardware

The hardware setup consisted of a CAVE with three video walls for stereoscopic visualization (passive stereoscopic projection, with a display resolution of 1400x1050; see Figure 11) and Infitec glasses and ART head tracking (Figure 12). Wavefield Synthesis System (WFS) for sound (Brandenburg, Brix, & Sporer, 2004; Pereira & Ebrahimi, 2002) due to its high-quality reproduction of the sound source positions when

compared to Dolby Digital 5.1 or 2-channel audio (stereo) was used. Users could hear the virtual agents as if they were in different places in the room—a parameter of the sound sources to the WFS renderer. The sound source parameters were extracted from the 3D-scene description, representing the speaker's voice.

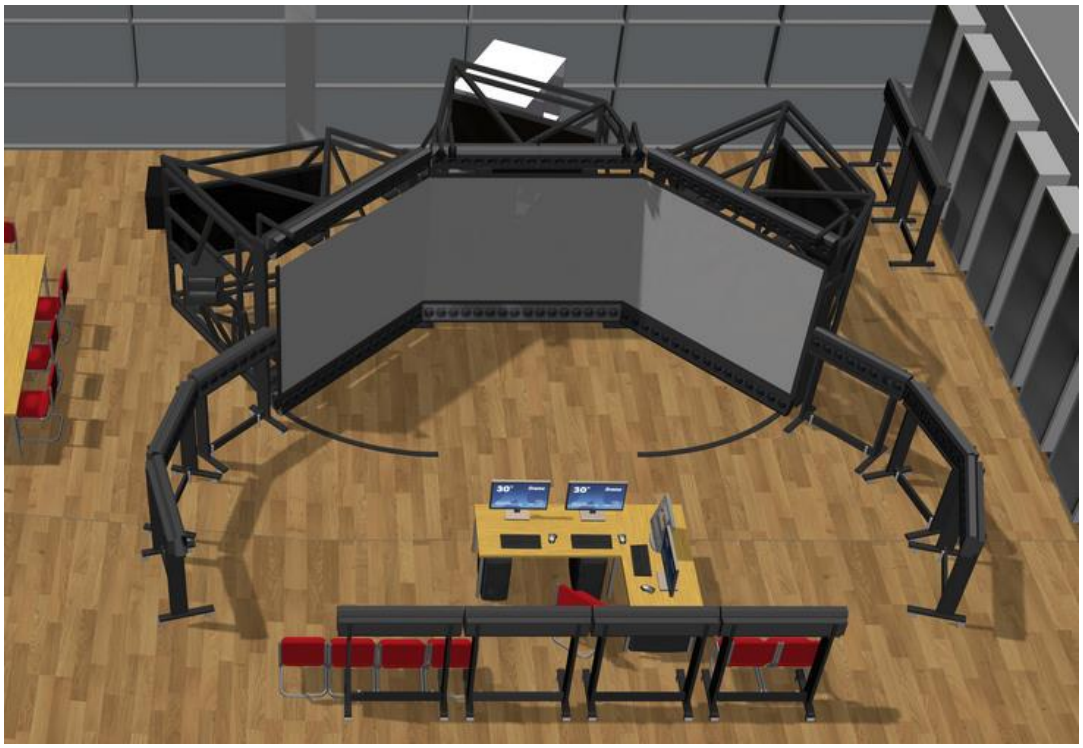

Figure 11: CAVE setup. Retrieved from https://www.tu-ilmenau.de/komwi/media-box/ab-medienproduktion-sose-2007/faszination-virtual-reality/. Copyright by Mariia Dubiago

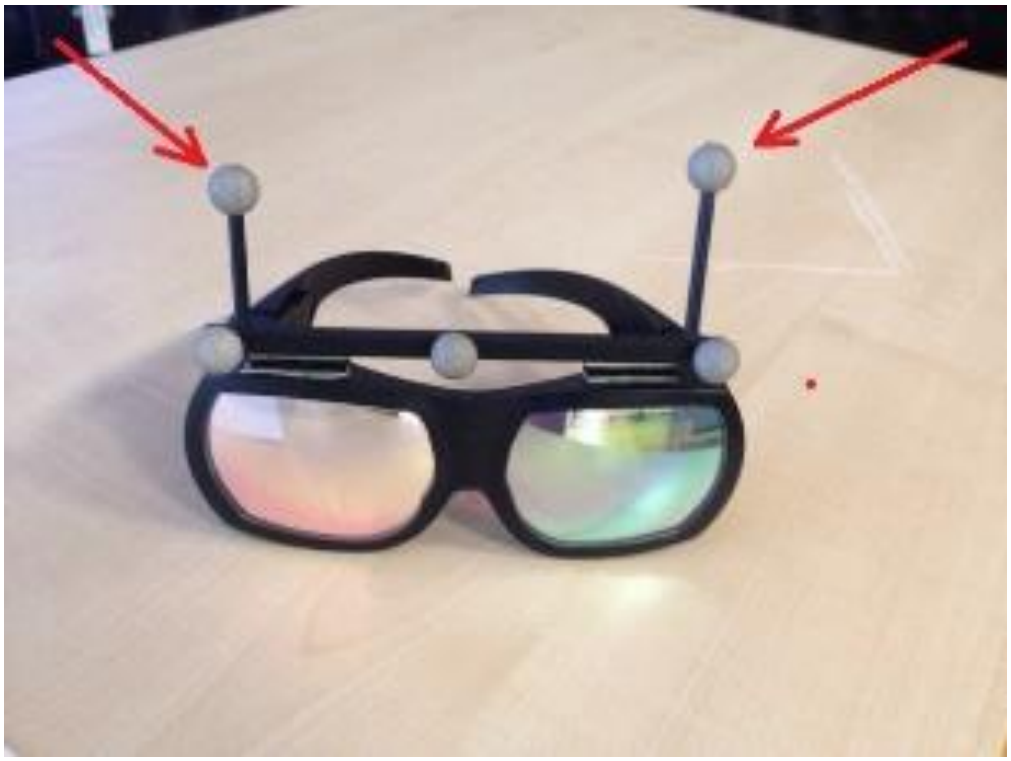

Figure 12: 3D glasses with tracking system. Own figure, 2016

The angle of the CAVE was adjusted to give users the feeling of being inside the classroom. The walls of the CAVE were closed for 90°, and the visual picture in the DeltaGen program, as well as a sound system, were adjusted accordingly. Figure 13 shows the CAVE adjustment to a 90° angle.

Figure 13: Screen setup. Own figure, 2016

3.3.2.3.Software

3D-Excite RTT DeltaGen 12.2 from the company Dassault for the description and rendering of the 3D-scene was used. Proprietary software was written by Dr. Drumm to describe 3D-scene behavior and to control audience behavior. It communicates with the integrated TCP-IP "External commands interface" of RTT DeltaGen 12.2 and receives necessary feedback information. This software controls the behavior of the audience as a mixture of random, predetermined, and manually controlled animations. Beyond that, it enables a visual interface (GUI) to be employed to vary a selection of parameters (e.g. repetition of animations) and supplies behavior statistics.

3.4. Results.

The results part contains findings from the implementation of the real speaker-audience interactive patterns into virtual reality public speaking training applications. The following paragraph describes the results of the study: development of new virtual characters (3.4.1), attentive, neutral, and non-attentive audience behavior (3.4.2), (manually controlled behavior (3.4.3), linguistic verbal communication (3.4.4), paralinguistic messages – intonations (3.4.5), non-linguistic communications – gestures (3.4.6), and non-linguistic communication – facial expressions (3.4.7).

3.4.1. Development of new virtual characters.

The virtual audience setup was built based on previous work and existed prototype developed at TU Ilmenau by Dr. Drumm. The existed prototype was lacking video data of certain behavior patterns needed for the interactivity (e.g., smiling facial expression), as well as behaviors assigned for attentive or non-attentive behaviors. The video data for existed virtual characters were developed by Arbitrarily Shape Video, recordings of real people, and thus, looked non-synthetic and realistic.

However, it created difficulties in extending the database, since it was not possible to recreate the exact recording environment with an identical look of actors (people who volunteered to be recorded for video data and be a virtual character's prototype in the application). Thus, seven new virtual characters were created. Seven real people were invited to be actors and be recorded for the video data. Developing new virtual characters involved three main steps: preparation of technical equipment as well

as a detailed description of the scenario, recording of a new video database, postproduction of recorded video data, and its incorporation into the application.

Preparation for New Characters Development. First, the list of linguistic, non-linguistic interactive behavior patterns for audience members was created. The list included a detailed description of all behavior patterns for each virtual character, including actions, postures, reactions, and interactions. All the behavior patterns were assigned to one of the three categories: attentive behavior, non-attentive behavior, and neutral behavior. The list of behavior patterns was created based on observation study and literature review on real audience behavior.

The existed application did not involve the attitude assigned behavior. Thus some of the behavior patterns were enlarged. For example, leaning right from existed application behavior list enlarged into leaning right with while nodding and keeping eye contact with the speaker and leaning right while having a bored facial expression. Table 32 shows examples of behavior patterns that existed in the VR public speaking prototype used as the basis for further developments and of behavior patterns added into the new prototype. Some utterly new behavior patterns were added as well (e.g. positive facial expression). The categories for the behavior patterns were adapted from existing VR public speaking application and extended for the research interest.

Table 32

Examples of behavior patterns existed in VR public speaking prototype used as basic and of behavior patterns added into a new prototype by categories

Category	Existed Prototype's Animation Example	New Prototype's Animation Examples
Postures	Lean back	1.Lean back while showing interest in the speaker 2.Leaning back while being bored and not looking at a speaker
Actions	Typing on computer	1.Typing notes on a computer while keeping eye contact with speaker time to time and nodding 2.Chatting on a computer, not keeping eye contact with the speaker and giggling
Reactions	No	Questions regarding the content of the presentation
Interactions	Audience members talk to each other about plans for the evening	Audience members talk to each other regarding the topic of the presentation Asking the speaker questions related to the topic

N = 7 participants (3 female, 4 male) voluntarily agreed to take part in the study and be recorded for the database that is used for creating new virtual characters. The light in the Media Lab was adjusted the way to avoid shadows. One 4k camera was placed in front of the table and facing the participant. With the analogy of existed video data, each virtual character had a starting position. With the starting position, each video from the database had to start and finish. The reason was to minimize glitches and "jumps" from one behavior being plied and other being started and, thus, make virtual audience behavior more realistic. The TV screen was used additionally to display the

starting position of the actor. The screen was seen by participants as well as their online video was overplayed on it. By that strategy used, each participant could adjust their position at the end of each behavior played matching to the starting one.

Recording Procedure. The recordings were conducted in the Media Lab of TU Ilmenau with the blue screen adjustments for recordings. Each participant was asked to sit in an upright position in front of the table in a Media lab setup with the blue screen behind. The table was covered with the blue textile as well to replace it later with the 3D table typically used in classrooms. One 4k camera was placed right in front of the participant. The microphone was placed in front of the speaker and synchronized with the video cameras. It allowed using sound data if it was needed.

The TV screen with starting position and online video played on it were placed on the right side from the recording camera, to make it valuable for participants. Three professional lights were used to project the light on the participant to avoid shadows. All the actions were recorded with a 4k video camera and saved on the computer. All participants gave informed consent. The setup for recordings is demonstrated in Figure 14.

Figure 14: Media Lab setup in TU Ilmenau for recording new virtual audience members' video database. Own picture, 2018

Postproduction. All the video data were post-processed. First, all the videos were cut for the right timeframe. Second, all the videos were rendered. Two types of rendering were done: for the alpha channel and color channel. The rendered images were combined back together for a complete video with MovieMaker. Both alpha and colored videos were added to a DeltGen program. For video that involved sound, the sound was extracted from microphone recordings, cut in the same timeframe as video, and added to a program, as well. As a result, each new virtual character had up to $N = 40$ behavior animations.

3.4.2. Attentive neutral and nonattentive audience behavior.

Additionally, three categories were used to define the audience's attitude: attentive behavior patterns, neutral behavior patterns, negative behavior patterns. Based on the literature review on nonverbal communication (Aldrich, 2004; Mehrabian, 1968; Poeschl & Doering, 2012a) all behavior patterns in the database were assigned into one of the three mentioned categories. Examples of three categories describing virtual audiences' attitudes are demonstrated in Table 33.

Table 33

Examples and descriptions of attentive, non-attentive and natural postures, actions, interactions, and reactions in new VR public speaking training prototype

Audience Design	Positive	Neutral	Negative
Postures	A neutral posture, sitting upright Leaning to different sides, attentive	Sitting upright while avoiding eye contact, looking around the room Leaning backward	Sitting upright, progressively moving to lean backward with arms folded Leaning (side) with arms folded, non-attentive posture
Behavior Actions	Mainly positive and neutral facial expressions Certain audience members taking notes with intermittent pauses to give eye contact to a presenter Certain audience members taking notes on a computer with intermittent pauses to give eye contact to a presenter	Neutral primarily facial expressions Audience members may be playing with pens or pencils to elicit a sense of disinterest Unrelated to speech behavior, like drinking water	Neutral and angry facial expressions Viewing direction are sparsely focused on the presenter Audience members may be twirling pens and drop pens Audience members use mobile phones or computers, without giving eye contact to a presenter

Table 33

Examples and descriptions of attentive, non-attentive and natural postures, actions, interactions, and reactions in new VR public speaking training prototype

Audience Design	Positive	Neutral	Negative
Interactions	Certain audience members take to engage in speaking to each other regarding the topic of presentation (e.g., discussing the nearly heard information)		Certain audience members take to engage in speaking to each other, potentially include sporadic audience laughter after a conversation
Reactions	Audience members may ask questions regarding the subject matter Audience members may give a positive reaction to the question's answer		Audience members may give a negative reaction to the answer to the question

3.4.3. Manually controlled behavior.

To create independent interactive virtual audience behavior, first, the possible interactive patterns were added into the application. As the virtual audience was able to play interactive behavior patterns, the new control panel was created to make the manipulation of the audience's behavior easy and fast. The control panel included six parts. First, virtual agents moving at the same time could be adjusted before starting the training, which influenced the activity of the virtual audience. By default, the number cannot be higher than six characters at the time, since it causes program overload and leads to errors. Second, it included the function of repetition/non-repetition of postures, actions, interactions, and reactions. That allowed the application user to adjust the

variety of user behavior patterns from the database. Third, the level of attentive behavior/neutral behavior/non-attentive behavior could be regulated in percentages online during the training. This allowed the operator to observe the speaker's behavior and change the audience's attitude accordingly. Fourth, it was possible to manipulate new characters to raise the hand. Fifth, all possible questions and reactions to speakers' answers were added to the control panel as buttons. So that operator could choose the question by choosing the virtual character and pressing the button with the keyword of the question. Finally, the audience could knock on the table and accepted as a form of polite applause in Germany. The Control panel for manual adjustment of virtual audience behavior is demonstrated in Figure 15.

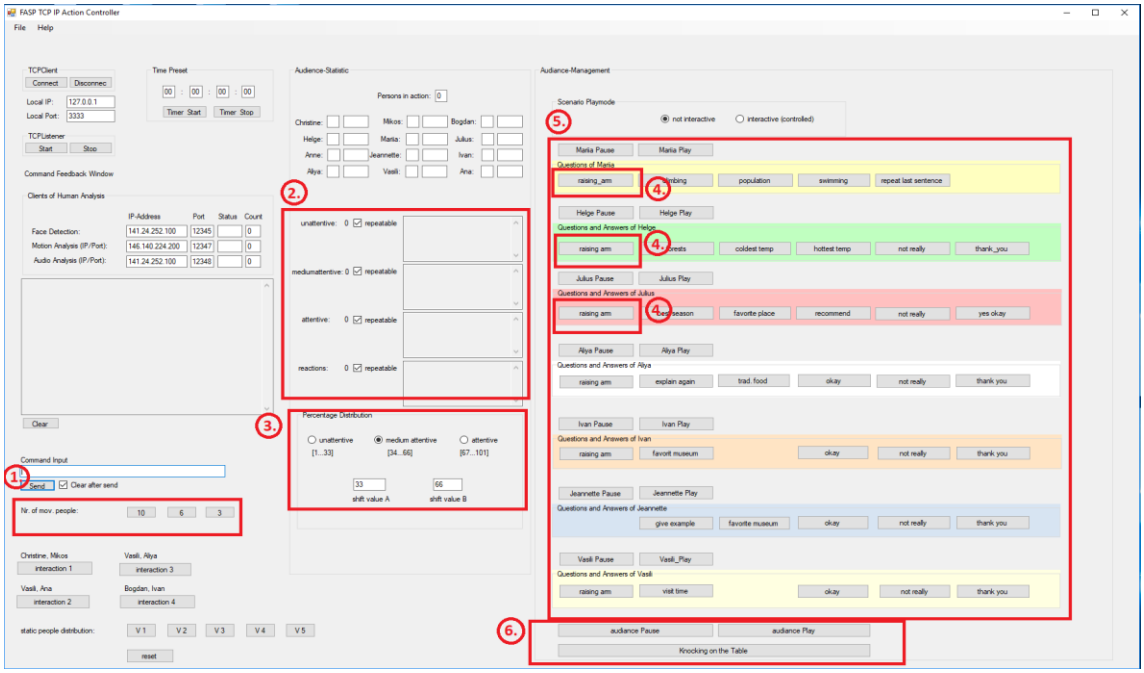

Figure 15: Control panel for manual control of virtual public speaking application designed in TU Ilmenau. Own picture, 2019.

3.4.4. Linguistic communication - verbal content.

The verbal speaker-audience interaction patterns discovered in the observation study were implemented in the prototype VR public speaking training application. For the verbal interaction, the manual control panel was created with the keywords for the questions.

Initiation of Dialogues. The virtual audience was manipulated to ask questions to the speaker only after such key sentences as ‘Do you have any questions?’ and not during the presentation, so that the application users (in the role of the presenter) were the initiators of dialogues with the virtual audience.

Addressees in Dialogues. There was always only one audience member having a dialogue with the user at the same time. Three audience members were asking questions related to the subject one by one so that the application user spent most of his/her dialogue time answering the questions from different audience members. The audience members were shown as always looking at the speaker during the answer and addressing questions to him/her.

Timing and Types of Dialogues. Dialogue appeared only towards the end of the scene after the speakers’ presentation. The virtual audience members mainly asked questions related to the subject matter of the presentations, and sometimes gave a reaction to the user’s answer (like ‘thank you for your answer’; Fig. 3). To implemented subjected-oriented questions, the topic of the presentations was limited to some presentations about different cities. Three audience members were able to ask three different questions due to the content (e g., “What was the coldest temperature in the city?”, “What is the population

of the city?"). The verbal reactions from the audience were generated by an operator based on the content of the presentation.

Technical Implementation in Application. To manipulate the application easier, the new interface for the application was created.

3.4.5. Paralinguistic massages – intonations.

Audience's Reactions and Speakers' Intonations. In the real audience, the overall voice volume was rated as medium to high. The virtual audience was manipulated the way, that when the volume of speakers' voice during the presentation was lower than average (as measured by five-voice recordings of different speakers) for longer than 30 seconds, one of the audience members ask the question "Can you please speak louder?". The own Audio Evaluation Tool for volume recognition was designed. Additional evaluation of the average volume of speakers was created.

The Audio Evaluation Tool was designed to recognize the volume of continuous speeches. The tool allows users to speak naturally, while the volume is being determined by the computer. Audio Evaluation Tool offers a special case of volume pattern recognition. Two phases were followed while developing the tool: training (recording speakers audio with volume evaluated as normal, low, and high) and testing. During the training phase, all the parameters of volume were recorded, and a pattern for low volume behavior was estimated using several case examples. During the testing phase, the feature of audio recognition was matched with the VR training application. The Audio Evaluation Tool was successfully implemented into VR application, and recognition of low volume behavior pattern was done, during the test phase. Frame 16 how the volume recognition tool looked during the audio recognition process.

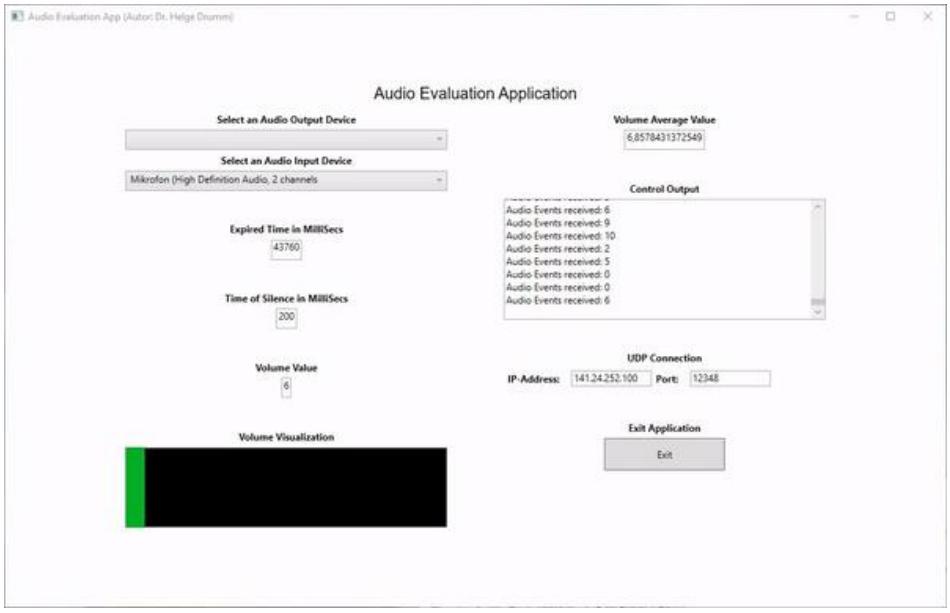

Figure 16: The own created Volume Recognition Tool. Own Figure, 2018

Audiences' Actions and Speakers' Intonations Only a significant association of interactions between audience members and the usage of intonations was found in real life. Recognizing intonation is a very complicated process that is very hard to implement. All existing systems for recognizing intonations at the moment have some drawbacks (Gibbon, Mertins, & Moore, 2000). There is still a question if the functional units of intonations can be identified, and if they can be identified without all the variation being accounted (Johns-Lewis, 1986). At the time when the interactive training application for public speaking was in implementation, there were no reliable tools for intonation recognition (including the accents and foreign speakers) exist. Thus, although there is a description of real audiences' reactions towards the speaker's intonations, the implementation of independent reactions towards certain intonations was not possible.

3.4.6. Non-linguistic communication - gestures patterns.

Frequency of Gestures used by Audience Members For the natural and regular behavior, not connected to attentive or non-attentive attitudes, such moves were used as body movement to front/back/side, head nods, open body posture, closed body posture. Each virtual audience member had a database of natural behavior postures and actions. The overall percentage of natural behavior used in the audience was adjusted according to observation study findings. The natural behaviors were performed by audience members randomly according to the manipulated distribution percentage.

Audience's Attitude and Speaker's Gestures Advanced technologies as Microsoft Kinect sensors were used to detect speakers' motions. The Microsoft Kinect sensor was placed in the virtual room on the table in front of the speaker, so it was able to recognize the speaker and his/her emotions through the presentation time. The recognition was done and analyzed in real-time. The skeleton of the recognized body and its motions can be observed in real-time from the recognition tool on the computer. The outlook of the gesture recognition tool can be seen in Figure 17.

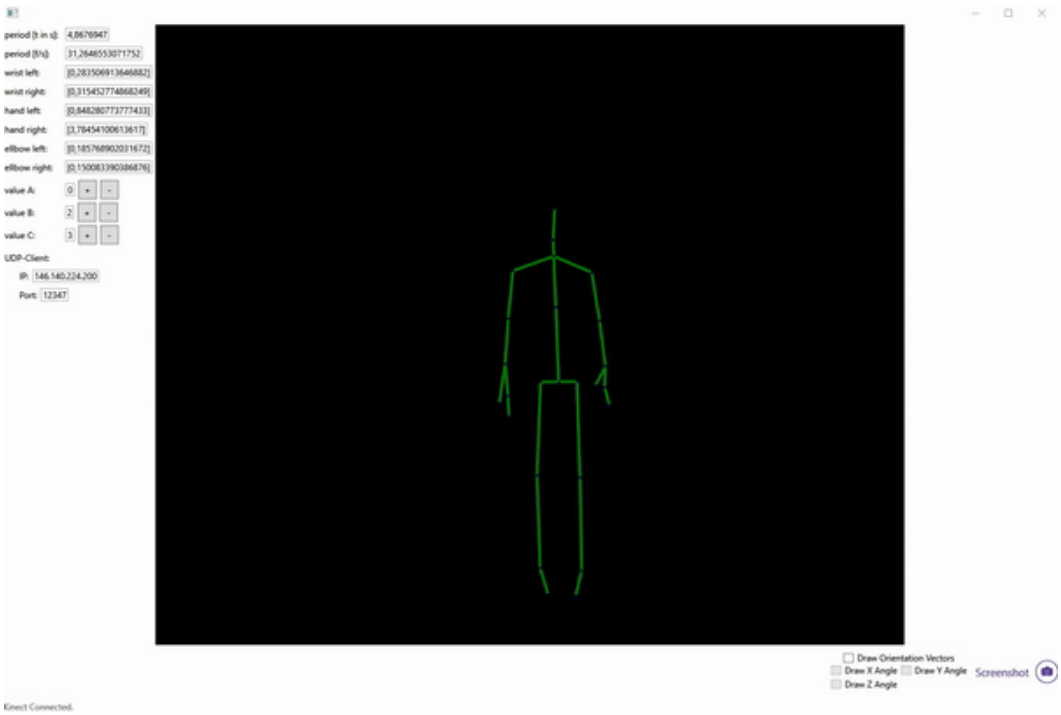

Figure 17: The motion recognition tool. Own Figure, 2018

The motion (gesture) recognition is the mathematical way of interpretation of human motions by a computing device; the tool used in this study was manipulated to measure the number of speakers' gestures. If the speaker is moving in a virtual classroom and using gestures while his/her speech, then the virtual audience members manipulated to display more postures and actions from the attentive behavior list. In case, if the speaker, does not use gestures and does not move while giving a speech, then the virtual audience is manipulated to play more postures and actions from the non-attentive behavior list. The percentage of speakers' movements directly influences the use of attentive or non-attentive behaviors by virtual audience members.

3.4.7. Non-linguistic communication - facial expressions.

Eye Contact. The virtual audience was manipulated to look at the speaker. Every user of a virtual training application is asked to wear the rim with tracking sensors on

the head during the speech. The tracking system aims to detect the positions and directions of the user and adjust the virtual space according to it. The tracing system was also used to simulate eye contact. Virtual audience members were following the speaker according to his/her location in the virtual space. Keeping eye contact with the speaker was assigned to a group of attentive behavior.

Looking at different objects or to other audience members were assigned to non-attentive behavior type and were performed by virtual characters as part of non-attentive behavior simulation.

Facial Expressions of Audience Members. The natural facial expression was used as a dominant facial expression of virtual audience members during the presentations. Additionally, the new virtual characters' positive, surprised, and negative facial expressions were added to a video database. Virtual audience members' positive facial expressions were used with only seven percent as a spontaneous facial expression towards speakers. Since such facial expressions as "negative" and "surprise" were used extremely rare by the real audience, they were not added as a virtual audience random behavior.

Often unclassified facial expressions used by the real audience were assigned to some actions and displayed as part of attentive/non-attentive/natural behavior parts. Facing down was assigned to looking at notes while writing action or other actions, including the natural need of looking in the direction of the table. Turning away was assigned to looking at other audience members while having a conversation or looking at some objects in the classroom while being non-attentive.

Audience's Facial Expressions and Speaker's Facial Expressions. To detect speakers' facial expressions, Affectiva emotion recognition was used. To be able to use the recognition tool, several changes in procedure had to be done. First, the used initially, 3D glasses were big for the program covered a big part of users' faces and made it impossible for a program to the face. Later, it was also found out that the use of Infitec 3D glasses of any sized was not possible.

Every Infitec glasses has a reflective coverage on top that makes a mirroring effect, which is distracting the recognition program. As a solution, polarized 3D glasses were used. This type of glasses has special lenses that can filter the image using the differences in light polarization for the right eye. Such glasses were tested, and they were successfully recognized by the recognition system. However, to use polarized glasses, a compatible projector that can polarize light differently for each eye was needed. For this purpose, a special lance for the projector with such ability was developed and attached to each of six projectors.

Finally, both the 3D system and emotional facial recognition were working successfully. The face and facial expressions of each user could be recognized and observed in real-time. Figure 18 demonstrates how the Affective face recognition tool looked during the recognition process.

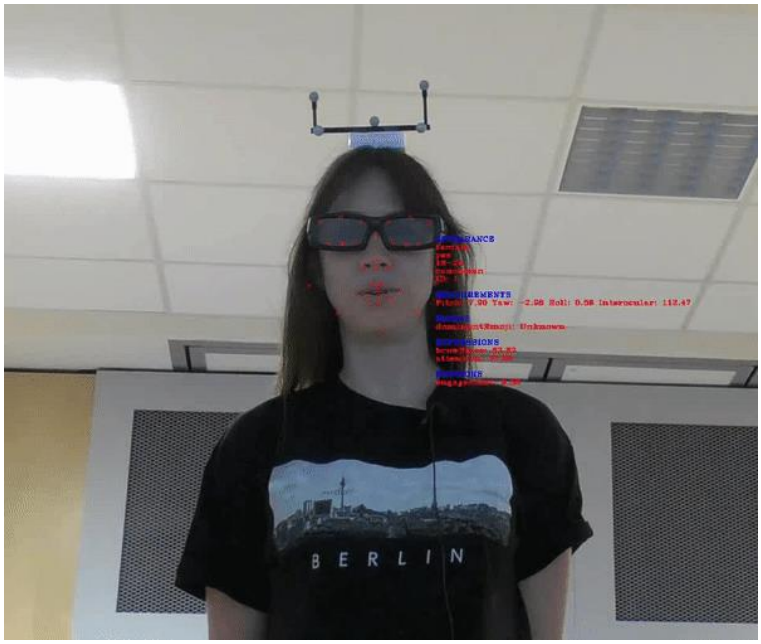

Figure 18: The face and emotion recognition tool. Own Figure, 2018

Finally, the virtual audience was manipulated to react to the speaker's smile. Whenever the users' smile was recognized, some random virtual audience members were manipulated to smile back.

3.5. Conclusions

There was no independent interactive virtual audience that existed at the moment. Developing virtual audiences with realistic behavior is one of the challenging tasks in the field (Goetz, Kiesler, & Powers, 2003; Slater, Sadagic, Usoh, & Schroeder, 2000). The second study of this dissertation thesis aimed to implement reactions of virtual audiences towards the speaker's behavior.

The second implementation study was a continuation of the observation study described earlier. The results obtained in the observation study show how people interact with each other during presentations in a seminar group, a mutually supportive

learning community. Those patters were used as speaker-audience interactive behavior and considered and patterns for implementation. Thus, the second implementation study was conducted. As a result of the study, several tools were used for interactive behavior implementation to implement reactions of the virtual audience towards the speaker's behavior. The following paragraph (3.5) summarizes all the findings of the implementation study (3.5.1), presents discussion (3.5.2), limitations, and strengths (3.5.3).

3.5.1. Summary of findings.

Realistic speaker-audience interactive behavior patterns were incorporated into a VR public speaking training application prototype. The findings from the user study indicated that interactivity of the virtual audience through speaker-audience dialogues, intonations, facial expressions, and gestures is highly relevant for future virtual public speaking applications.

Dialogues. The virtual audience was manipulated to ask questions to the speaker only after such key sentences as ‘Do you have any questions?’ so that the application users were the initiators of dialogues with the virtual audience. There was always only one audience member having a dialogue with the user at the same time. Three audience members were asking questions related to the subject one by one. Dialogues appeared only towards the end of the scene after the speakers’ presentation. The virtual audience members mainly asked questions related to the subject matter of the presentations, and sometimes gave a reaction to the user’s answer. To manipulate the application more comfortable, the new interface for the application was created.

Intonations. The Audio Evaluation Tool was designed to recognize the volume of continuous speeches. It allows users to speak naturally, while the volume is being determined by a computer. The virtual audience was manipulated the way that when the volume of the speaker was lower than average for longer than 30 seconds, one of the audience members ask the question, “Can you please speak louder?”. At the time that the interactive training application for public speaking was in implementation, there were no reliable and free available tools for intonation recognition exist. Thus, although there is a description of real audiences’ reactions towards the speaker’s intonations, the implementation of independent reactions towards certain intonations was not possible.

Gestures. Each virtual audience member had a database of natural behavior postures and actions that were performed by audience members randomly according to manipulated percentage distribution. Advanced technology as Microsoft Kinect was used to detect speakers’ motions. The motion (gesture) recognition is the mathematical way of interpretation of human motions by a computing device; the tool used in this study was manipulated to measure the number of speakers’ gestures. If the speaker is moving in a virtual classroom and using gestures while his/her speech, then the virtual audience members manipulated to display more postures and actions from the attentive behavior list. In case if the speaker does not use gestures and does not move while giving a speech, then the virtual audience is manipulated to play more postures and actions from the non-attentive behavior list. The percentage of speakers’ movements directly influences the use of attentive or non-attentive behaviors by virtual audience members.

Facial Expressions. The virtual audience was manipulated to keep eye contact with the speaker, and the natural facial expression was used as a dominant facial expression of virtual audience members most of the time during the presentation. Positive facial expression from audience members was added as random behavior and adjusted for 7%. To detect speakers' facial expressions, Affectiva emotion recognition was used. The virtual audience was manipulated to react to a speakers' smile and smile back.

All the findings from the implementation study structured by categories and findings from the observation study (Study 1) and summarized in Table 34.

Table 34

Summary of findings from implementation study structured by categories and findings from Study 1

Categories	Findings from Study 1	Implementation
Linguistic		
1. Initiator	Speaker	Initiation from speaker manually recognized
2. Addresser	One-to-one	Only one VR audience member participate in dialogue
3. Type	Questions	Q&A session with manual questions and answers generation

Table 34

Summary of findings from implementation study structured by categories and findings from Study 1

Categories	Findings from Study 1	Implementation
Paralinguistic Messages		
1. Question	Often used, leads to positive FE	Q&A session with a positive attitude from VR audience members
2. Incomplete	Often used, leads to neutral and positive FE	VR audience members, in general, have neutral and sometimes positive FE during the speech
3. Doubtful	Often used, lead to neutral and positive FE	VR audience members, in general, have neutral and sometimes positive FE as the during the speech
4. Confident	Not used	VR audience members, in general, have neutral and sometimes positive FE during the speech
5. Emphasized	Often used, lead to neutral and positive FE	The average speech volume level was measured
6. Volume	Not measured	No implementation
		VR audience members react to the speaker's low volume with the request to speak louder

Table 34

Summary of findings from implementation study structured by categories and findings from Study 1

Categories	Findings from Study 1	Implementation
Gestures		
1. Symbolic	Rarely uses	No implementation
2. Deictic	Rarely used	No implementation
3. Beat	Motor most used by the speaker, audience react with attentive behavior	Gestures recognition tool used to generate attentive or non-attentive virtual audience behavior VR audience members keep open body poster during the speech
4. Non-manual	Open body posture mostly used	
Facial Expressions		
1. Eye contact	Towards speaker	Eye contact with the speaker was simulated
2. Neutral	Mostly used	VR audience members most of the time have neutral FE
3. Positive	Used randomly and as a reaction to a speaker	VR audience members have positive FE occasionally during a speech; facial recognition tool is used to generate smiles in the audience as a reaction
4. Negative	Rarely used	No implementation
5. Surprised	Rarely used	No implementation

6. Unclassified	No effect	No implementation

Note. VR = Virtual Reality; Q&A = Questions and Answers; FE = Facial Expressions

3.5.2. Discussion.

VR technology helps to simulate real environments with a high degree of realism and interactivity (Bishop & Fuchs, 1992; Zeltzer, 1992), and VR training environments can expose users to a simulated situation in a safe, highly visual and interactive way (Aldrich, 2004; Ferry et al., 2004). Public speaking training applications is one of the contexts were VR technologies can be successfully used (Bodie, 2010; Furmark, Tillfors, Stattin, Ekselius, & Fredrikson, 2000) and to be successful, the applications should work on implementing speaker-audience interactions (Poeschl et al., 2014). Thus, this study aimed to implement realistic speaker-audience interactions in a virtual public speaking training application, and thereby to increase its quality.

To reach the main aim and answer the research question there are three steps done: (1) new interactive virtual characters were implemented, (2) interactive behavior was generated manually, (3) interactive behavior was implemented by implementing possible autonomous tools (Audio Recognition Tool, Microsoft Kinect, and Affective). Finally, the interactive speaker-audience behavior patterns obtained in the observation study were implemented into VR public speaking training application prototype using the possible autonomous recognition tools

As a result of the study, several interactive behavior patterns were implemented in the application. The speaker audience dialogues were implemented at first. The new interface for the application was developed the way that questions from virtual audience members could be generated manually. Secondly, the reactions towards a too low volume of users' speech have implemented the way that one of the virtual audience

members asked the following question: “Can you please speak louder, please”. Thirdly, the virtual audience was manipulated to perform more attentive behaviors towards speakers who use motions and gestures and all the way around. Finally, the virtual audience was manipulated to respond to a speaker’s positive facial expression with a smile.

It can be concluded that the first partly independent interactive virtual audience was developed. The virtual audience members with high visual appearance realism and behavior based on real audience behavior were developed. At the same time, the interaction between users and the virtual audience was implemented.

3.5.3. Limitations and Strengths.

The study has several limitations. One limitation of the study includes the fact that verbal communication, such as the questions from the audience and subsequent responses to the speakers’ answers, were still regulated manually by an operator. Thus, future studies should work on the atomization of the application by automating audience reactions.

Another limitation is that all the topic related questions from virtual audience members and topic related interactions between audience members addressed the same topic. Future work should increase the variety of possible speaker-audience dialogue forms and topics by using the more advanced (maybe paid) speech recognition tools that allow recognizing the topic of the speech.

Also, future work should have a possibility of virtual audience members generating automatically topic related questions with the voice generated tool. To implement the facial recognition tool, the older 3D technologies with worse resolution

were used. Future research should increase the quality of 3D technologies used in the project.

Finally, the study restricted technical equipment that was used limited to a particular setup that exists in TU Ilmenau. Further work should include more advanced technologies.

Despite these limitations, the study has several strengths. The newly designed VR audience was unique: While most of the existing applications use computer-generated virtual agents (Chollet, Ochs, & Pelachaud, 2014; Pertaub et al., 2002), in the research video-recordings of real people were employed to create an audience with a highly realistic appearance.

Additionally, the VR audience mostly behaved autonomously, in ways that were generated spontaneously and not regulated by an online operator or prepared in advance and related to the speaker's behavior during the presentation, except for the verbal interactions. Although the questions were still regulated by an operator, these dialogue patterns were the outcomes of a real audience behavior observation.

Generally, virtual audiences created by the use of advanced technical equipment and with highly realistic appearance and autonomous behavior lead to a high level of perceived realism (Bowman & McMahan, 2007b; Lee, Rincon, Meyer, Höllerer, & Bowman, 2013). For this reason, the presented VR public speaking application is of potential value in helping people to train under realistic conditions and reduce the fear of public speaking. Future work of this kind may use even more advanced technology, for example, 3D-records or live 3D-streams of real persons produced by 3D-depth cameras

(Beck & Froehlich, 2017). Another direction for future work is to implement the VR training application for a head-mounted display, such as Oculus Rift.

4. Evaluation Study

While the developers are trying to improve the efficiency of training applications by implementing interactive patterns, the question of how users perceive interactive applications remains open. Thus, the study aims to implement a newly created VR public speaking training application. The following chapter starts with the introduction and relevance of the evaluation study (4.1). Later, the theoretical background on effective VR public speaking applications is described (4.2). The chapter continues with the methodology of the evaluation study (4.3), results of the study (4.4), and conclusions of the study (4.5).

4.1. Introduction

77% of the general population suffers from a fear of public speaking (Hamilton, 2011). Fear of public speaking is one of the most common types of social phobia in the world (Lee et al., 2002b). Virtual reality (VR) public speaking training applications are a promising possibility to improve public speaking skills. However, to be effective and successful, VR training applications should evoke levels of state anxiety in VR that are comparable to those in real-life situations (Slater & Wilbur, 1997).

The development of high-quality VR public speaking applications demands a deep understanding of user-system interaction in VR and its determinants. The Quality Evaluation of User-System Interaction in Virtual Reality (QUEST-VR) framework (Poeschl, 2017) systemizes system features, user characteristics, system-user interaction, and moderating factors that determine the quality of VR applications. It provides a heuristic tool for evaluating interactive VR setups (Poeschl, 2017).

The framework includes such factors that influence the effectiveness of VR application as system fidelity (perceived realism), users' characteristics (as state and trait fear and confidence as a speaker), and moderating factor presence. Perceived realism is an "objective degree of exactness with which behaviors, rules, and object properties are reproduced in a simulation" (Ragan et al., 2015, p. 796). Fear (state anxiety) refers to fear in a specific situation, like giving a public speech (Menzel & Carrell, 1994). Speakers' confidence refers to the feeling or belief that one can make a speech performance (Hook, Smith, & Valentiner, 2008). Finally, presence "occurs when users feel that a form, behavior, or sensory experience indicates the presence of another individual" (Youngblut, 2003, p. 4). One way to influence these factors of application effectiveness is to implement realistic linguistic and non-linguistic speaker-audience interactions (Poeschl et al., 2014), i.e., dynamic sequences of actions between a human speaker and virtual audience members. Virtual audience members should modify their subsequent actions in response to their interaction partners (Turner, 1988).

The present work aims at evaluating how the implementation of speaker-audience interactive behavior in a prototype virtual public speaking training application affects user perception concerning perceived realism, feeling of fear, confidence as a speaker, and feeling of presence. Therefore, the third and last research question of this dissertation thesis was formulated:

RQ3: How does the interactivity of a virtual public speaking application affect user experience?

To answer these research questions, the experimental, cross-sectional study on perceived realism, the feeling of fear, confidence as a speaker, and the feeling of presence in the prototype was conducted.

4.2.Theoretical background on effective VR public speaking applications

Following paragraphs describe the state of research (4.2.1), the QUEST-VR framework (4.2.2) related to the evaluation study, as well as covers such vital topics as the role of perceived realism in VR applications (4.2.3), the role of the feeling of fear in VR applications (4.2.4), the role of confidence as a speaker in VR application (4.2.5), and the role of presence in VR applications (4.2.6).

4.2.1. State of research.

While developers work on technical implementations of virtual audience applications, scientists describe how people respond to virtual audiences. Slater et al. (1999) conducted a pilot experiment in which participants were giving a talk in front of the virtual audience displayed on the computer, but in a secluded and empty real room. It was stated that participants were affected by the behavior of the virtual audience.

Pertaub et al. (2002) continued the previous research and conducted an experimental study with three types of virtual audiences' behavior: positive, neutral, and negative. A significant correlation was found between the virtual audience behavior and participants' reaction. In particular, three types of virtual audience's behavior influenced the generated anxiety. Participants' response to virtual audiences with positive behavior was correlated positively with participants fear of public speaking (measured before the experiment). At the same time, participants' fear in response to the virtual audience with negative behavior was always high, independently from their fear of public speaking.

Thus, as stated by many studies, a virtual audience can elicit similar responses from participants as a real audience (Slater et al., 2006; Zanbaka et al., 2007). Such conclusions make virtual audiences applicable to different trainings. For instance, a virtual audience is used for musicians to train their performances in front of the audience (Bissonnette, Dubé, Provencher, & Moreno Sala, 2015).

Since participants respond to the virtual audiences in a way similar to the way how they respond to a real audience, new virtual applications were created and used in training. Public speaking skills can be trained, and fear of public speaking can be reduced in a safe but feared situation (Powers & Emmelkamp, 2008). VR has an advantage in such training as that people can role-play certain situations to overcome the fear in a safe virtual environment (Emmelkamp, 2013).

Researchers stated that VR training applications can be effective and can successfully reduce the fear of public speaking (Anderson, Rothbaum, & Hodges, 2003; Harris et al., 2002; North et al., 1997). North et al. (1997) made one of the very first studies that stated that VR can be successfully used in treating fear of public speaking. The self-reported level of fear was measured before and after the training. As a result, the reported fear level was reduced.

Later, Pertaub et al. (2002) conducted an experiment where he asked participants to make a speech in front of three types of different virtual audiences: neutral, negative, and positive. The results showed that the virtual audience behavior influenced participants, and all three virtual audience types generated a fear even for those participants who reported high confidence levels of public speaking.

Harris et al. (2002) conducted a study with students who had a noted level of public speaking fear. All students were divided into two groups. One group had training with a virtual audience, while the other group was a control group that has not had any training. The results supported the findings from the previous studies and demonstrated that VR training is effectively reducing the fear of public speaking.

Anderson et al. (2003) conducted an experiment where they asked participants with primary fear of public speaking to complete eight VR training sessions. The fear was measured before and after the trainings with self-reported measures. Participants' performance was rated by experts. The significant improvements in public speaking skills were made during the training sessions.

More research has been done later that supported the finding of VR exposer training effectiveness for improving public speaking skills and reducing fear of public speaking (e.g., Batrinca, Stratou, Shapiro, Morency, & Scherer, 2013).

Virtual audiences have shown potential in training applications. Public speaking training applications are a successful means to reduce public speaking fear symptoms and train respective social and presentational skills (Pertaub et al., 2002; Slater et al., 2006). Thus, researchers continue to study the effect of virtual training applications with virtual characters in it as well as looking for factors that can improve the outcome of such applications.

4.2.2. QUEST-VR Framework.

To be effective and successful, VR public speaking training application should feature users' feelings in a way that is comparable to real situations. The development of high-quality VR public speaking applications demands a deep understanding of user-

system interaction in VR and its determinants. The “Quality Evaluation of User-System Interaction in Virtual Reality (QUEST-VR) framework (Poeschl, 2017) systemizes system features, user characteristics, system-user interaction, and moderating factors that determine the quality of VR applications. It provides a heuristic tool for evaluating interactive VR setups (Poeschl, 2017).

According to the QUEST-VR framework, several features influence the outcome of the VR training applications and should be taken into account while designing the successful and effective VR application. First, the system itself is a crucial factor for the success of the application. It includes application context, task characteristics, and realism (LaViola, 2017). Developers can directly design these factors. In this study, the application context was developed to build on real-life observations. It also includes the fidelity factor, or *perceived realism* of the application that refers to the degree of exactness of VR environment and audience compared to the real-life experience (Ragan et al., 2015). Second, the user characteristics are a crucial factor and cover biological, physical, psychological, and social characteristics of users as well as the user’s limitations and capabilities that are needed for the specific training (Chapanis, 1991; Stramler, 1993). In the case of public speaking training applications, it can refer to *fear of public speaking* (Chaplin, John, & Goldberg, 1988) and to users’ *confidence as a speaker*.

Finally, QUEST-VR framework refers to moderating factors that influence the user-system interaction quality. One of the most research and controversial moderator factor is the *presence* (McCreery, Schrader, Krach, & Boone, 2013). All of the mentioned factors can influence the outcome of the VR application and, thus, are very

important for its success. Thus, the understanding of each factor is needed. The following paragraphs will describe the state of research for each factor in detail.

4.2.3. The role of perceived realism in virtual reality applications.

Even with the current developed technologies available, it is challenging to design and implement a realistic and believable virtual environment. However, VR applications attempt to replace real-life situations, and that is why they have sought maximum realism (Pujol-Tost, 2018). Traditionally, it was assumed that successful virtual applications and compared to real-life experiences of it are reached by developing the virtual application as realistic as possible (Fraser et al., 2000).

Several recent studies stated that the level of virtual training application realism affects feelings of presence and performance (Ragan, Wilkes, Cao, & Bowman, 2012). Herrlich et al. (2012) stated that a higher level of VR applications realism leads to a higher feeling of presence, to a higher level of performance, and a better transfer of gained skills into the real-life.

The realism of the application is, however, depending on many factors. First, the realism of the visual implementation is a crucial factor that influences the general perception of the VR application and even the feeling of presence (Hendrix & Barfield, 1996). The complexity and richness of the real world make the achievement of total visual realism of VR applications fundamentally tricky. Developers, trying to solve this issue by observing and imitating the colors, surface textures, reflections, graduations, and other details in the real environment (Foley, 1994). Allbeck and Badler (1998) stated that generating a virtual environment and virtual characters is possible.

On the other hand, it was also argued that training VR simulations with a low level of visual realism have enough information for a believable experience of a stressful situation, as it was successfully used in flight simulations (Meehan, Insko, Whitton, & Brooks, 2002). Moreover, the high level of visual realism will increase expectations for other realism factors, such as behavioral realism (Tromp et al., 1998). It can be concluded that the level of visual realism for the most effective VR application outcome is not clear, but it is critically important and does affect the outcomes of the application. That is why it should be measured and taken into account while developing successful VR applications.

The second factor that influences the general perception of VR application realism is connected to virtual humans in training applications. Lombard and Ditton (1997) defined VR realism as "the extent to which a medium can seem perceptual and/or socially realistic" (p. 187). On the one hand, the appearance of virtual characters is an essential factor that can influence the realism perception of the application in general as well as its effectiveness (Slater et al., 1999). Bailenson and Yee (2006) stated that "avatar realism is critical to the future of collaborative virtual environment development" (p. 670). Steed and Schroeder (2015) defined the realism of virtual audience appearance as one of the main factors that affect co-presence in VR.

On the other hand, the realistic behavior of virtual humans is also a crucial factor for the perceived realism of the application. In there are no changes in the virtual audience's behavior, or if the virtual audience is entirely static, participants will have a limited interest in the training procedure, no matter how well the virtual environment is designed (Cavazza et al., 2005). It was also stated in several studies that virtual audience

should show some nonverbal behavior towards the user to enable him/her to feel acceptable levels of presence while the exposure (Slater et al., 2000; Steed, Slater, Sadagic, & Tromp, 1999; Tromp et al., 1998).

Finally, the sound realism of the VR application plays a vital role in the general realism perception and increase the feeling of presence during the exposure (Larsson, Vastfjall, & Kleiner, 2002). It was proven that spatialized sound leads to a higher reported feeling of presence then non-spatialized sound or absence of sound at all (Hendrix & Barfield, 1996). Additionally, studies show that a combination of synchronized video and audio increases the sense of immersion (e.g., Larsson et al., 2002).

The concept of realism is quite broad and, as it was mentioned above, involves different aspects in it. Thus, it should be measured from different perspectives. This makes realism as a problematic factor to measure. However, it was stated that it is still the user who has to perceive and interpret the realism of provided Virtual Environment. Lee et al. (2013), for example, defined application realism as "the degree to which the images of the simulated world are perceived to be real by the user" (p. 548).

According to Witmer and Singer (1998), visual realism "does not require real-world content, but refers to the connectedness and continuity of the stimuli being experienced." (p. 232). That is why researchers state that the perceived realism of VR should be measured by using a self-report questionnaire. Poeschl and Doering (2013) proposed the German VR Simulation Realism Scale for measuring realism of the VR simulations with virtual humans in it. The German VR Simulation Realism Scale

includes four factors of realism evaluation that were based on the Witmer-Singer Presence questionnaire (Witmer & Singer, 1998).

The factors of the German VR Simulation Realism Scale included: (1) scene realism; (2) audience behavior; (3) audience appearance; (4) sound realism. The scene realism referred to the realism of visual cues such as colors, lights, reflections, three-dimensionality, and proportion realis. Audience behavior and audience appearance realism referred to virtual characters' realism. The audience behavior realism factor included such aspects as actions, postures, gestures, and facial expressions realism. Audience appearance realism was measured by including such aspects as the difference of virtual characters in appearance, the authenticity of the virtual humans for the situation, naturalness, and adequateness of virtual characters' outfit. The last factor referred to the realism of sound and ranked only one item (Poeschl & Doering, 2013). The German VR Simulation Realism Scale is the only scale that exists at the moment for measuring the perceived realism of virtual application with virtual characters in it.

4.2.4. The role of feeling fear in virtual reality applications.

There is growing evidence of the fact that people react to the VR audience in a similar way as to a real-life audience. Users with fear of public speaking report a similar level of fear while presenting in front of the virtual audience with fear while presenting in front of real audiences (Pertaub et al., 2002).

Pertaub et al. (2002) did research and measured participants' fear during presentations in front of virtual audiences with three different behavior types: positive, neutral, and negative. The findings of the study showed that participants who gave a speech in front of the virtual audience with negative behavior had a significantly higher

level of fear than participants who gave a speech in front of the virtual audience with the positive behavior. At the same time, it is stated that to be effective and successful, VR training applications should evoke levels of state anxiety in VR that are comparable to those in real-life situations (Slater & Wilbur, 1997). It can be concluded that the VR training application for public speaking should evoke fear, and the feeling of fear should be comparable to the fear evoked in real-life public speaking situations.

Short et al. (1976) stated that perceived effectiveness is more important than actual effectiveness. It is the same with the feelings. Thus, feeling of fear during the VR training can also be measured by self-reported measurements (Wiederhold & Wiederhold, 2005b).

There are several scales exist to measure fear by a self-reported questionnaire. One of the first scales developed for self-reported anxiety measurements is the Cognitive-Somatic Anxiety Questionnaire (CSAQ) by Delmonte and Ryan (1983). It consisted of 14 items in total, seven for somatic, and seven for cognitive indicators. Other scales were developed shortly, including improved and more validated items. The Worry-Emotionality Scale (WES) developed by Morris et al. (1981) included ten indicators with five items for worry (cognitive concern) and five items for emotional (somatic arousal). Lecrubier et al. (2000) created Lehrer-Woolfolk Anxiety Symptom Questionnaire with 27 items for all behavioral, cognitive, and somatic indicators.

The scale that was most recently developed and mostly used in nowadays research is a Shortened State and Trait Anxiety Inventory (SSTAI) by Marteau and Bekker (1992a). It is an updated, and shortened version of the STAI scale by Spielberger

(1983) included originally 20 indicators. The SSTAI was shortened to finally six items (the reliability and validity for the scale were kept).

4.2.5. The role or confidence as a speaker in VR applications.

The research stated that the negative behavior of the VR audience triggers fear irrespective of participants' confidence in public speaking (Pertaub et al., 2001). Slater et al. (1999) asked all the participants to complete the Personal Report of Confidence as a Speaker (PRCS; Hook et al., 2008) questionnaire designed to assess the confidence as public speakers before experiencing the VR public speaking exposure. They stated that the experiment was conducted with participants who had relatively high levels of public speaking confidence. As the outcome of the study was the fact that negative virtual audiences triggered state anxiety, although the participants had a normal level of public speaking confidence. However, the influence of virtual audience behavior on the speaker's confidence was not measured.

Slater et al. (2006) conducted an experiment where they asked confident speakers and participants with a fear of public speaking to present in an empty room and in front of a virtual audience. They concluded that a confident speaker had the same response while speaking to an empty room and while speaking to a virtual audience. People with a fear of public speaking had significantly different trends of heart rate from confident speakers while presenting in front of a virtual audience. However, researchers claimed that the virtual characters had low representational quality, and the audience behavior was pre-programmed and independent from users' behavior.

The confidence of the speakers in VR public speaking training applications was not well researched. Most of the current research analyzes the influence of VR training

application on fear, but there is none of the research was done for analyzing the VR training influence on speakers' confidence as a speaker. At the same time, the speaker's confidence is a crucial factor for public speaking experience and can be influenced by the feared stimulus as well as it is inevitable an important outcome of the public speaking training.

There are two main ways to assess individual differences in confidence are exist in studies, those are (1) self-report questionnaires that are designed to find out one's belief in his/her ability to perform a public speaking task; and (2) judgments of likelihood of success, after the completion of a public speaking task (Stankov, Kleitman, & Jackson, 2015). There are few scales exist with the specific aim of measuring confidence during public speaking, such as Report of Communication Apprehension (McCroskey, 2006), Self-statements During Public Speaking (Hofmann & Dibartolo, 2000), or the Personal Report of Confidence as Speaker (PRCS; Gilkinson, 1942; Paul, 1967). Of these three measures, one of the most frequently used is the PRCS.

The original PRCS scale was a self-report one with 104 true or false items and was developed to measure a participants' fear of public speaking. Participants were asked to feel the questionnaire based on their most recent public speaking experience. However, the original scale proved to be cumbersome. Later, a short version with a 30-item scale was developed. Thirty most discriminating items were selected from the original scale. Again, the questions were designed in a format of true or false answers with the total score rating from 0 – false to all items to 30 – true to all the items. However, the scale has limited psychometric data. Also, there was only one study conducted to assess the structural validity of the scale (Hook et al., 2008).

Later, Bartholomay and Houlihan (2016) proposed The Public Speaking Anxiety Scale (PSAS; Bartholomay & Houlihan, 2016) that contains 17 items with a five-point Likert scale. The scale was derived from several previous public speaking anxiety scales, considered as empirically-based psychological assessments. It has a high internal consistency ($\alpha = .938$; Bartholomay & Houlihan, 2016, p. 214) PSAS is a highly reliable tool with a valid measure to assess cognition, behaviors of speech anxiety (Bartholomay & Houlihan, 2016). Another advantage of the PSAS is that it is much shorter than other existed scales. At the same time, the length of the questionnaire is considered a big concern in the experiment studies when participants need to spend much time in the laboratory.

4.2.6. The role of presence in virtual reality applications.

Sheridan (1992) used the term "presence" for a journal-title that published studies related to VR systems, and that is how the term entered the general scientific discussion. The term was used in a wide range of mediums, such as television and telephone (Lombard & Ditton, 1997), films (Zeltzer, 1992), and others. Lee and Nass (2005) applied term of presence to the whole mediated experience in general and pointed out that "presence is at the heart of humans' desire to use media to move beyond the limits of body and the sensory channels" (p. 3). In the sphere of VR, the term presence recently received much attention. VR has a higher quality than the other media and, thus, a higher level of presentation quality that VR applications make possible (Biocca, 1997).

The first definition of presence was given by Lombard and Ditton (1997), and it was phrased as "the illusion of non-mediation." This short definition was later often

repeated in other studies. VR public speaking training applications to be effective should evoke the feeling of ‘being there’, or feeling of presence (Slater & Wilbur, 1997). Barfield and Weghorst (2009) stated the importance of the feeling of presence for VR applications and mentioned that “presence in a virtual environment necessitates a belief that the participant no longer inhabits the physical space but now occupies the computer-generated virtual environment as 'place'” (p. 289). A bit later, the presence was defined as “being in another place other than the one you are physically in” (Schroeder, 2002, p. 3). The (International Society for Presence Research, 2000) gave a broader definition to presence:

> *“a psychological state or subjective perception in which even though part or all of an individual’s current experience is generated by and/or filtered through human-made technology, part or all of the individual’s perception fails to accurately acknowledge the role of the technology in the experience. Except in the most extreme cases, the individual can indicate correctly that s/he is using the technology, but at some level and to some degree, her/his perceptions overlook that knowledge and objects, events, entities, and environments are perceived as if the technology was not involved in the experience.”*

The feeling of presence is significant in designing successful VR applications because it allows people to behave in VR in the same way as they would behave in real-life (Slater et al., 1994). Thus, training VR applications aim to evolve users feeling of presence during the training, and presence is considered as a critical factor in VR

training applications (Wiederhold & Wiederhold, 2005a). The studies proved the importance of presence and its significant correlation with self-reported anxiety level (Emmelkamp, Bruynzeel, Drost, & van der Mast, 2001; Price & Anderson, 2007).

The study by Ling et al. (2014) conducted a meta-analysis with 33 articles and 52 correlations and could support the findings. There was a moderate association between self-reported feelings of presence and anxiety during VR exposure therapies for anxiety disorders. It was concluded, that feeling of presence in VR training applications plays an essential role and influence the success of the training (Wiederhold & Wiederhold, 2005a). It was stated that to be successful and effective, VR public speaking training applications should evoke the feeling of 'being there' or feeling of presence (Slater & Wilbur, 1997).

> *"The participants who are highly present should experience the V.E as more the engaging reality than the surrounding physical world, and consider the environment specified by the displays as places visited rather than as images seen. The behaviors in the V.E should be consistent with behaviors that would have occurred in everyday reality in similar circumstances"* (p. 166).

To measure the feeling of presence, there are several questionnaires designed. However, most of them are focused on a place of illusion or virtual presence (Slater, 2009). The feeling of presence is a more brought term that involves more aspects. Lombard and Ditton (1997) differentiated spatial and social presence, while Ijsselsteijn et al. (2001) differentiated physical presence and social presence. However, most of the

studies stated that the concept of feeling of presence in VR applications covers three main parts: social presence, co-presence, and spatial presence (IJsselsteijn, Bouwhuis, & Ridder, 2004; Lombard et al., 2000).

Social presence is defined as "the moment by moment awareness of the co-presence of another, being accompanied by a sense of engagement with the other (i.e., human, animate or artificial being), the sense of being together" (Biocca, Harms, & Gregg, 2001, p. 1) or as "the subjective experience of being together with others in a computer-generated environment, even when participants are physically situated in different sites" (Youngblut, 2003, p. 4). Social presence refers to the feeling of social interaction with other beings, including virtual humans and to the feeling of being together with other beings (Heeter, 1992; IJsselsteijn et al., 2004; Lombard & Ditton, 1997). Thus, social presence is highly relevant for VR training applications involving virtual humans, avatars, or virtual audiences.

Co-presence is defined as the ability of VR users to perceive people in the VR environment and estimate how the user thinks that the audience perceived him/her (Goffman, 1963). Goffman (1963) stated that "co-presence renders persons uniquely accessible, available, and subject to one another" (p. 22). Finally, spatial presence defined as "the sense of being in the virtual environment rather than in the environment in which they are physically located" (Witmer & Singer, 1998, p. 226). Spatial presence refers to the physical sense of being placed in the virtual environment and not in the real environment (Heeter, 1992; Lombard & Ditton, 1997).

It is still challenging for researchers to measure the feeling of presence because it is based on subjective feelings of the VR users. There are several scales for

measurements that were proposed, and most of them were in the format of a self-reported questionnaire (Youngblut, 2003). Several questionnaires were often used in different studies. For example, such as Igroup Presence Questionnaire (IPQ) by Schubert et al. (2001); Slater-Usoh-Steed Questionnaire (SUS) by Usoh et al. (2000); Presence Questionnaire (PQ) by Witmer and Singer (1998); and Independent TelevisionCommission (ITC) - Sense of Presence Inventory (ITC-SOPI) by Lessiter et al. (2001).

Nevertheless, all of the mentioned questionnaires include only items that are measuring spatial presence. Poeschl and Doering (2015) developed the German Scale for a Fear of Public Speaking Scenario to measure co-presence and social presence, especially in public speaking VR training situations. The developed scale was based on items from existing questionnaires and items extracted from the state of research. The German Scale for a Fear of Public Speaking Scenario included four subscales and measured “the presenter’s reaction to virtual agents, perceived virtual agents’ reaction to the presenter, impression of interaction possibilities, and presence of other people in the virtual space (co-presence)” (Poeschl & Doering, 2015, p. 62).

4.3.Methodology

The study was conducted in the Flexible Audio-visual Stereo Projection system (FASP) of the Ilmenau University of Technology. The study evaluated VR public speaking training prototype. The training application was developed by recording videos

of 12 people and creating a database of postures, actions, reactions, and interactions for each virtual character. The virtual environment was created with RTT DeltaGen 12.2.

The unique environment of VR public speaking training application with several advantages was developed at TU Ilmenau and was used for this study. The virtual seminar room was matching the actual seminar room in TU Ilmenau. Each virtual character in the virtual classroom was able to change his/her posture, display different actions, reactions, and interactions. The virtual audience consisted of seven male and five female characters sitting in a classroom, with desks oriented in order with three lines and four rows. In this study, two types of the virtual audience were developed: an interactive virtual audience that reacted to the speaker's behavior, and non-interactive audience that behaved independently from speakers' behaviors.

The application was running on a PC. Participants were navigated into the virtual environment using a tracking system and 3D glasses. The position of the virtual classroom was adjusted before participants arrived and corrected according to participants' height when participants arrived at the laboratory, by the use of wand in the CAVE and by the use of the mouse on the computer. Participants were asked to stand in front of the virtual audience, were wearing 3D glasses and tracking sensors.

The methodology and research design of the study at hand is explained in this chapter. In the following, the method, sample, and construction of the codebook are presented. Data collection and data analysis processes are described thereafter.

The following paragraphs describe the research question of the evaluation study (4.3.1), questionnaire development (4.3.2), sample (4.3.3), data collection and content analyses (4.3.4), design and data analyses (4.3.5), and ethical consideration (4.3.6).

4.3.1. Research question.

The goal of the study was to reveal how participants perceive interactive virtual audience for public speaking training application. With this goal, the third research question, as mentioned above, was formulated as:

RQ3: How does the interactivity of a virtual public speaking application affect user experience?

For this study, a quantitative method was employed to obtain a current evaluation of the application. VR applications attempt to replace real-life situations, and that is why it has sought maximum realism (Pujol-Tost, 2018). Traditionally, it is assumed that successful virtual applications and compared to real-life experiences of it are reached by developing the virtual application as realistic as possible (Fraser et al., 2000). Interactivity is a big part of the communication process that includes the audience passive involvement in perceiving and understanding the speech and giving linguistic and non-linguistic reactions. Interactive behavior is a natural process that happens in real audiences, thus, it was hypnotized that improvement of interactivity of virtual audience can improve the perceived realism of the application and the first hypothesis was formulated as:

H1: Higher interactivity of the application leads to a higher perceived realism.

People react to the VR audience in a similar way as to a real-life audience. It was stated that participants who speak in front of the virtual audience with negative behavior had a significantly higher level of fear than participants who give a speech in front of the virtual audience with the positive behavior. The behavior of the virtual audience positively influences the feeling of fear. Since interactivity will make the virtual

audience behavior more natural and closer to real audience behavior, the second hypothesis was formulated as:

H2: Higher interactivity of the application leads to a higher feeling of fear.

Following the same logic, if interactivity will lead to a higher feeling of presence; consequently, it will reduce the confidence of the speaker during the presentation. Thus, the third hypothesis was formulated as:

H3: Higher interactivity of the application leads to lower confidence of the speaker.

Presence in VR training applications plays an essential role and influence the success of the training. However, it provides a general situation and lacking personal links that are needed for an increase in the feeling of presence (Wiederhold & Wiederhold, 2005a). The interactive audience provides personal feedback that depends on a particular speaker's behavior. Thus, it is expected that an interactive audience will increase the feeling of presence. The fourth hypothesis was formulated as:

H4: Higher interactivity of the application leads to a higher feeling of social presence.

4.3.2. Questionnaire development.

Perceived realism: Perceived realism was assessed using the German VR Simulation Realism Scale (Poeschl & Doering, 2013). The questionnaire consists of 13 (α = .812) items that constitute the three sub-scales scene realism (five items; e.g., reflection, light, and shades in virtual space), audience behavior realism (four items; e.g., postures and gestures of virtual humans), and audience appearance realism (four items; e.g., outfit of virtual humans). All perceived realism evaluation items were based

on a five-point Likert-scale ranging from 1 “highly disagree” to 5 “highly agree”. Table 35 demonstrates all sub-scale items.

Table 35

Sub-Scale Items for Perceived Realism From the German VR Simulation Realism Scale (Poeschl & Doering, 2013)

Sub-Scale	Item
Scene Realism	Reflection in virtual space seemed to be natural.
	Light and shades in virtual space were realistic.
	The virtual space seemed to be three-dimensional.
	Coloring in the CAVE appeared to be natural.
	Proportions of the virtual space were realistic.
Audience Behavior	The posture of virtual humans was natural.
	Gestures of virtual humans were natural.
	The behavior of virtual humans in the CAVE was authentic.
	Facial expressions of virtual humans were realistic.
Audience	Outfit of virtual humans was adequate.
Appearance	Virtual humans differed concerning their appearance.
	Virtual humans in their entirety seemed to be authentic for this occasion.
	Outfit of virtual humans was natural.
Sound Realism	Please evaluate the sound in the virtual room.

Feeling of Fear: The shortened version of STAI was incorporated in a survey to measure the feeling of public speaking fear during the presentation in front of the virtual audience (state anxiety). Self-reported measurements are still considered as the most reliable and easier available tool to measure and assess state anxiety during the VR simulation. The STAI short form included six items ($\alpha = .742$) as I feel calm; I feel tense; I feel upset; I am relaxed; I am worried; I feel content. All feeling if fear

evaluation items were based on a Likert scale 1 “totally disagrees” option and 4 “totally agree”.

Confidence as a speaker: The Public Speaking Anxiety Scale (PSAS; Bartholomay & Houlihan, 2016), which included 17 items (α = .843) with a five-point Likert scale. The value ranged from 1 “totally disagree”, to 5, “totally agree”. This scale was chosen because first, it derives from several previous PSA scale with high internal consistency; Bartholomay & Houlihan, 2016, p. 214). All the items that were included in The Public Speaking Anxiety Scale are summarized in Table 36.

Table 36

Items from The Public Speaking Anxiety Scale From The Public Speaking Anxiety Scale (Bartholomay & Houlihan, 2016)

1. Giving a speech is terrifying.
2. I am afraid that I will be at a loss for words while speaking.
3. I am nervous that I will embarrass myself in front of the audience.
4. If I make a mistake in my speech, I am unable to re-focus.
5. I am worried that my audience will think I am a bad speaker.
6. I am focused on what I am saying during my speech*.
7. I am confident when I give a speech*.
8. I feel satisfied after giving a speech*.
9. My hands shake when I give a speech.
10. I feel sick before speaking in front of a group.
11. I feel tense before giving a speech.
12. I fidget before speaking.
13. My heart pounds when I give a speech.
14. I sweat during my speech.
15. My voice trembles when I give a speech.
16. I feel relaxed while giving a speech*.
17. I do have problems making eye contact with my audience.

Note: reversed items

Presence: Effective VR training applications should evoke the feeling of presence. Hence, this research is interested in measuring if the new VR training

application with a higher level of interactivity allows participants to feel a higher feeling of presence. Often it referred to subjective experience (Ijsselsteijn et al., 2001; Sheridan, 1992; Witmer & Singer, 1998). Thus, subjective reporting is one of the most used tools to measure the presence (de Greef & IJsselsteijn, 2000).

In this research to measure the feeling of presence during the presentation in front of the virtual audience, the Presence and Social Presence in Virtual Environments Scale developed by Poeschl and Doering (2015) were used. The scale included three factors to identify the feeling of presence. In total, 15 items ($\alpha = .714$) were used to identify the feeling of presence. The five-point Likert scale from 1 “highly agree” to 5 “highly disagree” was used for all the factors. All the factors and items are summarized in Table 37.

Table 37

Factors and Items from the Presence and Social Presence in Virtual Environments Scale (Poeschl & Doering, 2015)

Sub-Scale	Item
Presenter's Reaction to Virtual Agents	People's behavior influenced my style of presentation.
	People's behavior influenced my mood.
	I reacted to people's behavior.
	I was easily distracted by the people.
Perceived Virtual Agents' Reaction	Sometimes the people were influenced by my mood.
	Sometimes the people were influenced by my style of presentation.
	The people reacted to my actions.
	I was able to interpret people's reactions.
The impression of Interaction Possibilities	I had the feeling to interact with other human beings.
	I felt connected to other people.
	I had the feeling that I was able to interact with people in the virtual room.
	I had the impression that the audience noticed me in the virtual room.
(Co-)Presence of other people	I was aware that other people were with me in the virtual room.
	I had the feeling that I perceived other people in the virtual room.
	I felt alone in the virtual environment.

4.3.3. Sample.

Firstly, to test the difference between two independent groups means using a two-tailed test, a priori power analysis was conducted using G*Power3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) with a medium effect size (d = 0.15), and an alpha of 0.05. The result showed that for achieving a power of .80, a total sample of 86 participants with two equal-sized groups of n = 43 was required. However, according to

Alroobaea and Mayhew (2014 - 2014) for analysis of the user experience and usability studies, 20 participants per group gain much validity in user testing.

Participants (N = 82) from a German university, who gave informed consent, were recruited for the study via university mailing lists and personal contacts. Due to missing data for three participants, additionally, n = 22 participants were excluded due to the technical irritations during a presentation, presenting in other languages rather than English, or giving presentation off-topic. The final sample consisted of N = 57 (n = 31 for the non-interactive condition; n = 26 for the interactive condition) subjects (65% male, 35% female, Mage = 25.98 years, SD = 4.68 years).

4.3.4. Data collection.

To collect the data, the qualitative method was employed. Participants filled pre-questionnaire regarding the trait anxiety and demographic data, then experienced the program and later asked for their open thoughts and subjective feelings through the questionnaire with close-ended, rating questions. The questionnaire consisted of four parts that pertained to the research objectives. The independent variable was manipulated by randomly assigning participants to interactive or non-interactive virtual audience behavior, to examine changes in presenters' perception of VR application.

To answer the research question, four variables as a perceived realism, a feeling of fear, confidence as a speaker, and feeling of presence were treated as equal dependent variables. Manipulated stimuli as two types of virtual audience behavior were treated as an independent variable.

Upon invitation to the study, participants were asked to prepare a five-minute presentation on a self-selected subject without the use of any electronic devices. They

were also asked not to read the whole presentation merely, but to speak freely instead. Through these restrictions, users were forced to look at the simulated audience. Participants were informed upon invitation that they are supposed to give a speech to a virtual audience in English.

After arriving at the laboratory, subjects were briefed and gave a speech in the virtual public speaking training application. Before and after the speech, participants filled in the questionnaire. German VR Simulation Realism Scale (Poeschl & Doering, 2013), Public Speaking Anxiety Scale (Bartholomay & Houlihan, 2016), The Spielberger State-Trait Anxiety Inventory (STAI) Y-6-item short form (Marteau & Bekker, 1992b), Presence and Social Presence in Virtual Environments (Poeschl & Doering, 2015).

4.3.5. Design and data analyses.

The goal of the study was to reveal how the interactivity of the VR application affects the participants' perception of the public speaking training application. To reach this goal, the VR setup was prepared in advance. An experimental, cross-sectional study was conducted. The variables measured in this research had a subjective sense. Thus, a subjective measurement approach was applied (Sheridan, 1992).

Two types of VR public speaking application prototypes were developed and pre-tested. The first VR prototype included a non-interactive virtual audience: a virtual audience was developed without recognition tools, and its behavior was predefined to a natural type of behavior. The behavior of a non-interactive virtual audience did not depend on speakers' behavior. The second VR prototype included an interactive virtual audience: a virtual audience was developed with the use of different recognition tools

described earlier, and the audience's behavior was not predefined. The behavior of interactive virtual audience depended on speakers' behavior during the presentation (question and answers session at the end of the presentation, reaction to speakers' speech volume, reaction to gestures intensively, and reaction to speakers' facial expression).

An independent measure, the between-subject study design was used. All the participants were randomly assigned to experience one of two VR public speaking prototypes. Before experiencing one of the VR prototypes, all the participants were asked to fill the pre-questionnaire. Immediately after using one of the VR training application prototypes, in which participants experienced interactive or non-interactive VR audience behavior, they were asked to evaluate the VR experience. There are four dependent variables measured: perceived realism, the feeling of social presence, a feeling of fear, and a feeling of confidence. There is one independent variable – interactivity – that consists of two independent categories, namely "interactive" versus "non-interactive" virtual audience behavior.

The questionnaires were then subjected to quantitative content analysis. A one-way within-subjects multivariate analysis of variance was run to determine the effect of virtual audience interactivity on application perception. Four measures of application perception were assessed: perceived realism, feeling of social presence, feeling of fear, and feeling of confidence. There is one independent variable – interactivity – that consists of two independent categories, namely "interactive" versus "non-interactive" virtual audience behavior. All analyses were performed with IBM SPSS Statistics 22.

The means and standard deviations for the four dependent variables are presented in Table 38.

Table 38

Mean Scores and Standard Deviations for Measures of Perceived Realism, Feeling of Fear, Confidence, and Feeling of Social Presence

	Perceived Realism		Feeling of Fear		Feeling of Confidence		Feeling of Social Presence	
Virtual Audience Behavior	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Interactive	3.60	.49	41.92	10.29	3.66	.68	3.54	.41
Non-Interactive	3.81	.38	39.46	9.19	3.83	.58	3.30	.42
Total	3.72	.44	40.58	9.70	3.75	.63	3.41	.43

Notes. $N = 57$.

Before running the MANOVA, all of its statistical assumptions were tested. According to Field et al. (2012), there are four assumptions of the test that should be considered. Firstly, observations have to be statistically independent, which means that there is no relationship between the observations in each group of the independent variable or between the groups themselves. This was given by having different participants in each group. Secondly, the data has to be randomly sampled and measured at an interval level. Thirdly, the MANOVA needs the data to be multivariate normal. It is expected that in MANOVA, the dependent variables "have multivariate normality

within groups" (Field et al., 2012, p. 717). Multivariate normality is a particularly tricky assumption to test for and cannot be directly tested in SPSS Statistics (Bray & Maxwell, 1985). Multivariate normality of the data was investigated by using information from multivariate skewness and kurtosis using the SPSS test of normality, which provides skewness, kurtosis, and Shapiro-Wilk's test.

Perceived realism, feeling of fear, and feeling of presence was normally distributed as assessed by Shapiro-Wilk's test ($p > .05$), but not for confidence as a speaker. However, since the sample sizes were greater than 50, graphical methods of analyzing such as a Normal Q-Q Plot was used to check multivariate normality for confidence as a speaker, since at larger sample sizes the Shapiro-Wilk test is highlighting as statistically significant even minor deviations from normality (Field, 2009). Confidence as a speaker was normally distributed for both conditions, as assessed by Normal Q-Q Plot.

Finally, there should be homogeneity of variance-covariance matrices. A further assumption of the one-way MANOVA is that there are similar variances for each dependent variable (Field et al., 2012, p. 717). This assumption can be tested using Box's M test of equality of covariance. The variance-covariance matrices were homogeneous, as assessed by Box's M test of equality of covariance matrices ($p = .823$) and by Levene's Test of Homogeneity of Variance ($p_{\text{Persieved Realism}} = .249$, $p_{\text{Feeling of Fear}} = .469$, $p_{\text{Feeling of Confidence}} = .393$, $p_{\text{Feling of Presence}} = .642$).

After testing the statistical assumptions, the data was checked for practical issues, that is, multicollinearity and outliers. First, the dependent variables were tested for multicollinearity by running a series of Pearson correlations that were pooled across

the two groups (Meyers, Gamst, & Guarino, 2013). The general rule of thumb is that if the simple correlation coefficient between two variables is higher than .80 or .90, multicollinearity is a severe problem (Grewal, Cote, & Baumgartner, 2004). However, according to Tabachnick and Fidell (2013), this is only a problem if the variables correlate highly positively. Highly negative correlations, on the other hand, are ideal for a MANOVA.

Table 39 shows the results of the Pearson correlation tests. The lowest correlation was observed between a feeling of confidence and a feeling of fear ($r = –.55$, $p < .001$). The highest correlation was observed between perceived realism and feeling of presence ($r = .31$, $p = .019$). The tests obtained a meaningful pattern of correlations between the dependent variables. Thus, it was concluded that there is no multicollinearity.

Table 39

Mean, Standard Deviations, and Correlations Among and Descriptive Statistics for Four Dependent Variables

Variables	Perceived Realism	Feeling of Presence	Feeling of Fear	Confidence as a Speaker
Perceived Realism		.31*	–.15	.25
Feeling of Presence			.15	–32*
Feeling of Fear				–.55**
Feeling of Confidence				

Note. $N = 57$. For perceived realism, feeling of presence and confidence as a speaker 0 = negative for all items, 5 = positive for all items. For a feeling of fear 0 = not realistic, 40 = realistic for all items,

* $p < .05$,

** $p < .001$

Secondly, the data was checked for *univariate outliers* in any of the dependent variables. This was achieved by boxplots, one of the most used methods for examining the presence of outliers. The boxplots were checked for values greater than 1.5 box-lengths from the edge of the box. There were no univariate outliers in the data, as assessed by inspection of a boxplot for values higher than three box-lengths from the edge of the box. Preliminary assumption checking revealed that data was normally distributed. Thus, the data were subjected to one-way within-subject multivariate analyses.

4.3.6. Ethical consideration.

The experimental study was conducted with the ethics principals of the American Psychological Association (Smith, 2003). The study was partially conducted together with the Master Human-Computer Interaction in Immersive Environments seminar class at TU Ilmenau in the summer semester of 2018. Students participated in the preparation of study materials such as questionnaires, invitation letters, consent forms, and in the organization of the experiments, including the pretest. All the participants were informed about the data confidentiality policy and signed the informed consent. No personal data was inquired. The researchers' contact information was provided for any questions. All the participants were rewarded with incentives such as USB sticks.

4.4.Results

A one-way within-subjects multivariate analysis of variance (MANOVA) was run to determine the effect of virtual audience interactivity on application perception. Four measures of application perception were assessed: perceived realism, feeling of social presence, feeling of fear, and feeling of confidence. There is one independent variable – interactivity – that consists of two independent categories, namely "interactive" versus "non-interactive" virtual audience behavior.

Preliminary assumption checking revealed that data was normally distributed, as assessed by Shapiro-Wilk test (Perceived realism: $p_{\text{interactive}} = .829$, $p_{\text{non-interactive}} = .138$; Feeling of fear: $p_{\text{interactive}} = .413$, $p_{\text{non-interactive}} = .350$; Confidence as a speaker: $p_{\text{interactive}} = .060$, $p_{\text{non-interactive}} = .160$; Fear of presence: $p_{\text{interactive}} = .488$, $p_{\text{non-interactive}} = .680$) and by Q-Q Plot examination. There were no univariate or multivariate

outliers, as assessed by boxplot and Mahalanobis distance ($p > .001$), respectively; there were linear relationships, as assessed by scatterplot.

The MANOVA revealed a statistically significant difference between interactive and non-interactive virtual audiences on the combined dependent variables, $F(4, 52) = 3.25$, $p = .019$; Wilks' $\Lambda = .800$; partial $\eta^2 = .200$. Table 40 shows detailed results for the multivariate and univariate analysis of variance for two audience conditions (interactive and non-interactive).

Only one independent variable showed a statistically significant difference between the two conditions. The feeling presence in interactive VR was evaluated significantly higher ($M = 3.54$, $SD = 0.41$) than the social presence of a non-interactive virtual audience ($M = 3.30$, $SD = 0.42$; $p = .034$, $\eta^2 = .079$). Perceived realism in the non-interactive virtual audience was higher ($M = 3.81$, $SD = 0.38$) then the realism in the interactive virtual audience ($M = 3.61$, $SD = 0.49$), which meant that the covariation was not going in the hypothesized direction. Participants' fear in the interactive virtual audience was higher ($M = 41.92$, $SD = 10.29$) then the fear in the non-interactive virtual audience ($M = 39.46$, $SD = 9.19$), so the covariation was going in the hypothesized direction. Finally, the confidence of the speaker was higher during a performance in front of the interactive virtual audience ($M = 3.66$, $SD = 0.58$) than during performance in front of the non-interactive virtual audience ($M = 3.83$, $SD = 0.58$), and the covariation was going in the hypothesized direction.

Table 40

Multivariate and Univariate Analysis of Variance for Two Audience Conditions

Group	Multivariate			Univariate											
				Perceived Realism			Feeling of Fear			Confidence as a Speaker			Feeling of Presence		
	F^a	p	η^2	F^b	p	η^2	F^b	p	η^2	F^b	p	η^2	F^b	p	η^2
Audience Type	3.25	.019	.200	3.01	.089	.052	0.91	.345	.016	1.08	.303	.019	4.72	.034	.079

Note. $N = 57$.

Since the covariation of the feeling of realism was not going in the hypothesized direction, the general evaluation of perceived realism was checked as it is presented in Table 41, the feeling if realism was evaluated as rather high.

Table 41

Frequencies and Percentages of Evaluated Feeling of Realism

Feeling of Realism	Frequency	Percentage
Low	2	3.5
Middle	13	22.8
High	42	73.7
Total	57	100

4.5.Conclusions

The evaluation study is the third and last study of this dissertation thesis. The evaluation study used the VR public speaking training application developed in the implementation study and interactivity of which, as implemented based on the first observational study. The evaluation study aimed to find out how the implementation of speaker-audience interactive patterns in a VR public speaking training application affects user perception concerning perceived realism, feeling of fear, confidence as a speaker, and feeling of presence.

To reach the research goal, the experimental, cross-sectional study on perceived realism, feeling of fear, confidence as a speaker, and feeling of presence in the VR public speaking training application was conducted. The summary of findings,

discussion, limitations, and strengths of the evaluation study presented in the following paragraphs.

The following paragraphs summarize the findings of the evaluation study (4.5.1), presents the discussion (4.5.2), limitations, and strengths (4.5.3).

4.5.1. Summary of findings.

The goal of the study was to reveal how participants perceive interactive virtual audience for public speaking training application, and the research question was formulated as:

RQ3: How does the interactivity of a virtual public speaking application affect user experience?

A one-way multivariate analysis of variance was run to determine the effect of virtual audience interactivity on application perception. Four measures of application perception were assessed: perceived realism, social presence, fear, and confidence as a speaker. Participants experienced two different conditions: interactive virtual audience behavior and non-interactive virtual audience behavior. Preliminary assumption checking revealed that data was normally distributed, as assessed by Shapiro-Wilk tests ($p > .05$) and by Normal Q-Q Plot examination. There were no univariate or multivariate outliers, as assessed by boxplot and Mahalanobis distance ($p > .001$), respectively; there were linear relationships, as assessed by scatterplot; no multicollinearity; variance-covariance matrices were homogeneous, as assessed by Box's M test ($p = .798$).

The first hypothesis was not supported. Perceived realism in the non-interactive virtual audience was higher ($M = 3.81$, $SD = 0.38$) then the realism in the interactive virtual audience ($M = 3.61$, $SD = 0.49$), and the difference was not statistically

significant (p = 0.089). Frequencies for the realism of the virtual public speaking prototype was evaluated rather high in general for both conditions (73.7%), and the most of participants (68.3%) did not have VR experience before the participation in the experiment.

The difference between interactive and non-interactive conditions was going in the expected direction, and interactive virtual audience feeling of fear was higher (M = 41.92, SD = 10.29) in a non-interactive virtual audience (M = 39.46, SD = 9.19). However, the result was not statistically significant.

The difference between interactive and non-interactive conditions was also going in the expected direction, and interactive virtual audience confidence of the speaker was higher (M = 3.66, SD = 0.58) then the confidence of the speaker during a performance in front of the non-interactive virtual audience (M = 3.83, SD = 0.58). However, the difference was not statistically significant.

As expected, the hypothesis was supported, and participants' feeling of presence (M = 3.54, SD = 0.41) in the interactive virtual audience was significantly higher than the feeling of presence in the non-interactive virtual audience (M = 3.30, SD = 0.42; p = .034, η^2 = .079).

There was a statistically significant difference between interactive and non-interactive virtual audiences on the combined dependent variables, $F(4, 52) = 3.25$, p = .019; Wilks' Λ = .800; partial η^2 = .200

4.5.2. Discussion.

This study was conducted in FASP TU Ilmenau, and it was carried out to evaluate how the interactivity of the new VR public speaking training prototype that was

created in the implementation study (study 2) influences users' perception. The first test of the interactive VR public speaking training application showed positive feedback as well as evoked high perceived realism, feeling of fear, and feeling of presence. There were two conditions of VR application developed: the interactive virtual audience that reacted to the speaker's behavior during the presentation and non-interactive virtual audience that behaved independently from the speaker's behavior. The experimental cross-sectional study with $N = 57$ participants was conducted, where participants were randomly assigned to one of two conditions. Four hypotheses were formulated to reach the goal.

The MANOVA analysis was conducted to find the difference between the two conditions. It provided general information of how four dependent variables as perceived realism, feeling of fear, confidence as a speaker, and feeling of presence affected by different interactivity level of the virtual audience. At the same time, descriptive statistical analyzes have given the information on how the realism of the application was evaluated by participants in general.

Considering the data obtained from the quantitative research, it can be concluded that the stimuli – interactive and non-interactive virtual audience behavior – affect users' perception.

Successful virtual applications are reached by developing the virtual application as realistic as possible (Fraser et al., 2000). The first hypothesis was formulated as:

H1: Higher interactivity of the application leads to a higher perceived realism.

However, the hypothesis was not supported. Perceived realism in a non-interactive virtual audience was higher than the realism in interactive virtual audience.

Additionally, frequencies for the evaluated feeling of realism were checked. It was noticed that the realism of the virtual public speaking prototype was evaluated rather high in general for both conditions (73.7%). At the same time, most of the participants did not have VR experience (68.3%). Thus, it can be concluded that due to the lack of VR experience, participants evaluated both interactive and non-interactive VR public speaking training prototype equally as highly realistic. Further research should examine participants who have experience in VR use and can identify differenced more carefully as well as to be more precise in their examinations.

People tend to react towards the VR audience similar to a real-life audience. Since interactivity will make the virtual audience behavior more natural and closer to a real audience behavior the second hypothesis was expected as:

H2: Higher interactivity of the application leads to a higher feeling of fear.

Even though the difference between interactive and non-interactive conditions was going in the expected direction and interactive virtual audience feeling of fear was higher in a non-interactive virtual audience, the result was not statistically significant. The study was conducted with participants not suffering from high degrees of fear of public speaking. Therefore, the feeling of fear was not high on average among all the participants. Also, it remains an open question of how phobic participants will react to the VR public speaking application.

Following the same logic, it was expected that interactivity would lead to a higher feeling of presence. Consequently, it will reduce the confidence of the speaker while the presentation and the third hypothesis was formulated as:

H3: Higher interactivity of the application leads to lower confidence of the speaker.

The difference between interactive and non-interactive conditions was going in the expected direction, and interactive virtual audience confidence of the speaker was higher than the confidence of the speaker during a performance in front of the non-interactive virtual audience. However, the difference was not statistically significant. Participants of this study most likely, have prior experience of presenting in precisely such situations, namely in front of casually looking, young people. Thus, they could feel confident during the presentation in front of the virtual reality public speaking prototype. Hence, future audiences should be created and evaluated for a broader range of target groups (e.g., politicians).

The feeling of presence in VR training applications influences the efficiency and outcome of the training (Wiederhold & Wiederhold, 2005a). The state of research shows that a high level of presence leads to performance closer to real-life experience, and a better transfer of gained skills into real-life (Kothgassner et al., 2012b). It was expected that the interactive audience would increase the feeling of presence, and the fourth hypothesis was formulated as:

H4: Higher interactivity of the application leads to a higher feeling of social presence.

As expected, the hypothesis was supported, and participants' feeling of presence in interactive virtual audiences was significantly higher than the feeling of presence in a non-interactive virtual audience. It means that the independent grouping variable (interactive and non-interactive VR audience) simultaneously explains a statistically

significant amount of variance in participant's feeling of presence. It explains that there is a statistically significant difference in the presence feeling between interactive and non-interactive conditions.

A higher feeling of presence leads to a higher level of performance, and a better transfer of gained skills into the real-life (Herrlich, Malaka, and Masuch, 2012). Thus, there is a potential that interactive VR public speaking training prototype allows participants to use VR training applications in a way similar to a real-life experience as well as allows better participants' transfer of gained skills to a real-life situation.

Three out of four hypotheses were not supported. Nevertheless, in the case of a relatively small sample, it is normal that observed large differences in means between two groups might not be significant. Thus, future studies should include a bigger sample for reviling more differences between the two conditions. Still, reactions of the presenters toward the interactive and non-interactive virtual audience appear to be relevant with medium effect sizes. This can be explained by the fact that interaction between people is an integral part of any kind of social behavior, including public speaking situations.

4.5.3. Limitations and strengths.

The study has limitations that could potentially influence the outcomes and offer opportunities for further research. First, there was a methodological limitation, such as that the sample size was quite small. As a result, it was challenging to identify the difference between the two groups. Hence, the differences in means were quite small. Second, the convenient international sample with no native English speakers was used. At the same time, the application was developed only in the English language. Some

questions that form a virtual audience could not be clearly understood. Future studies should include possibilities for the application in different languages or include native English speakers into the sample. Thirdly, the study was conducted with participants not suffering from very high degrees of fear of public speaking. Most of the participants had prior experience of presenting in precisely such situations, namely in front of casually looking, young people. Thus, they could have evaluated this application as very realistic, and it remains an open question of how phobic participants will react to the VR public speaking application. Forth, all participants were randomly assigned to one of two conditions. However, it was not checked if the manipulation was recognized. Further studies should let participants experience both conditions and see if the manipulation is being recognized. Additionally, future studies can include more target groups to improve the VR training application for a broader range of possible user groups (e.g., politicians, school pupils, businesspeople).

Despite the limitations, the study had its strengths. First, the self-report scale that covered indicators for individual evaluation was used that gave a clear idea on users' perceptions. Second, the first evaluation study for the interactive audience was conducted, and the study covered four most discussed topics in a field of VR public speaking training applications as perceived realism, feeling of fear, confidence as a speaker, and feeling of presence. Finally, the study contributes towards a better understanding of the user's perception of interactive VR public speaking training application compere to non-interactive ones.

5. General Discussion

This dissertation thesis contains three studies. First, the interactive behavior of the real speaker and the audience were examined, and speaker-audience interactive behavior patterns were summarized. Second, the speaker-audience interactive patterns were implemented into VR public speaking training applications. Finally, the interactive VR public speaking training application was evaluated. The summary of all the findings from all three studies is summarized in paragraph 5.1. Later, the general limitations and strengths of the dissertation thesis summarized in paragraph 5.2.

5.1. Summary of findings

To hold a speech in front of an audience is an essential skill in both private and professional life (at least in many occupations). However, the majority of the population lacks public speaking skills (Hart, Gratch, & Marsella, 2017) and tends to feel stressed before or during public talks. The leading psychological training method for fear of public speaking is cognitive-behavioral training, where people are systematically exposed to fear-triggering stimuli (Heimberg & Becker, 2002). As an outcome of this training, fear of public speaking can be reduced, and public speaking skills can be significantly improved (Harris, Kemmerling, & North, 2002).

However, real human audiences are not always available for logistical reasons, or a real human audience can be too intimidating for phobic trainees even to start the training. New technologies, such as VR, can provide the opportunity to expose users to

simulated situations safely and interactively (Aldrich, 2004) and can be successfully used in training public speaking skills.

To foster the transfer of trained skills into practice, VR training applications should feature users' feelings similar to real situations (e.g., feeling of presence, fear; Bishop & Fuchs, 1992). One way to influence user perception of virtual public speaking applications is to implement realistic speaker-audience interactive behavior (Poeschl, Tudor, & Doering, 2014). However, the ways through which humans gain social skills and how technology can maximize this process are an under-researched domain in human-computer interaction (Slovák & Fitzpatrick, 2015). At the same time, little research on the behaviors of real human audience members and speaker-audience interactions exists (Poeschl et al., 2014; Tudor, Mustatea, Poeschl, & Doering, 2014).

The dissertation thesis aimed to *develop and evaluate a realistic and interactive audience for the VR public speaking training application*. To reach this goal, firstly, an observation study on real speaker-audience interactions was conducted. The patterns identified in this study were implemented in the prototype application (implementation study). Finally, the quantitative evaluation study of the new interactive prototype was conducted to identify users' perceptions.

Observation Study. The research aimed at implementing realistic speaker-audience interactive behavior patterns in a virtual public speaking application prototype. Due to the lack of data on real human *speaker-audience interaction behavior patterns*, the observational study was conducted.

Structured, non-participant overt observation study on speaker-audience interactive behavior during real presentations in an English-language master program

class at a middle-size university in Germany was conducted. Speakers, as well as audiences, were video-recorded with four video cameras to capture different angles. The participants comprised $N = 14$ (nine female, five male) audience members and $N = 8$ different speakers (four female, four males), all of whom had given informed consent. Event sampling was used to determine the frequency of speaker-audience dialogues and speakers' intonations. Time sampling was used to determine the frequency of facial expressions and gestures. The sample resulted in a $N = 6{,}484$ interactive behavior patterns that were subjected to quantitative content analysis.

The results obtained in the observation study showed how the speaker and audience interact with each other during presentations in a seminar group, a mutually supportive learning community. It was found that speakers are the more active partners in the dialogues, speakers and audience members often create questions to interact with each other. Further, the audience, most of the time, keeps eye contact with the speaker and natural facial expression during the presentation. Thirdly, the audience reacts to speakers' positive facial expressions and gestures. Although the convenient international sample with no native English speakers was used for the study and both audience and speakers were familiar with the settings, the first study analyzing interactions between real audience and the real speaker was successfully conducted. Based on the results of the study, all the interactive speaker-audience behavior patterns were summarized and structured following the social behavior theory. Also, for each of the findings, there were recommendations for further implementation into the VR training application described. Study 1 provides a clear structure of speaker-audience interactive behavior

patterns that can be used in creating realistic interactive VR public speaking training applications. This information is highly relevant and closer to the exciting research gap.

Implementation Study. The ways of *implementing speaker-audience interactive patterns* identified in the observation study had to be found next. The hardware setup consisted of a CAVE with three video walls for stereoscopic visualization (passive stereoscopic projection, with a display resolution of 1400x1050), Infitec glasses, and ART head tracking. Wavefield Synthesis System (WFS) was used for sound so that users could hear the virtual agents as if they were in different places in the virtual classroom. The software setup was realized with 3D-Excite RTT DeltaGen 12.2. Proprietary software was written to describe the 3D-scene behavior and to control audience behavior. It communicated with the integrated TCP-IP "External commands interface" of RTT DeltaGen 12.2 and received necessary feedback information. This software controls the behavior of the audience as a mixture of random, automatic, predetermined, and manually controlled animations.

Following findings from observation study, manual adjustments were used to implement speaker-audience dialogues, the virtual audience was manipulated to ask questions to the speaker only after such key sentences as 'Do you have any questions?' and asked questions related to the subject matter of the presentations. General settings of virtual audience behavior, such as eye contact and facial expressions, were adjusted. Interactive behavior was developed by using free tools.

A self-created audio recognition tool was used to recognize speakers' volume levels. The virtual audience was manipulated to react on a low volume level and request to speak louder. Kinect motion recognition tool was used for recognizing speakers'

gestures and movements. The virtual audience was manipulated to behave attentive or non-attentive, depending on the number of speakers' gestures and motions. Affectiva facial expression recognition was used for facial recognition. The virtual audience was manipulated to react to the speaker's smile and smile back (a few random virtual agents from the audience).

Interactive behavior patterns were successfully implemented into VR public speaking training applications. However, verbal reactions from the virtual audience were still regulated manually. Despite this, new interactive VR public speaking training application was successfully created and included unique interactive speaker-audience behavior patters based on observations of interactive speaker-audience behavior in a real-life situation.

Evaluation Study. Evaluation of whether the implementation of speaker-audience interactive behavior patterns met *user expectations* was done. A quantitative user study to evaluate the VR public speaking prototype concerning the perceived feeling of social presence, fear, confidence as a speaker, and realism was conducted (based on QUEST-VR framework; Poeschl, 2017). Participants were acquired through a university mailing list and oral invitations. The final ad-hoc sample (N = 57) consisted of 37 men (65%) and 20 women (35%; M = 26.0, SD = 4.7). All participants gave informed consent. Participants were asked to prepare a five-minute presentation about any city in the world, without the use of any electronic devices.

After arriving at the laboratory, subjects were briefed and gave a speech in the virtual public speaking training application. All subjects were randomly assigned to whether new interactive virtual audience (including implemented interactive behavior

patterns as described in implementation study) or to a non-interactive virtual audience (audience with random behavior not connected to speakers' behavior). After the speech, participants filled in Presence and Social Presence in Virtual Environments Scale (Poeschl & Doering, 2015), The Personal Report of Confidence as a Speaker. Public Speaking Anxiety Scale (Bartholomay & Houlihan, 2016), Communication Anxiety Inventory (Marteau & Bekker, 1992), and the German VR Simulation Realism Scale (Poeschl & Doering, 2013).

A one-way multivariate analysis of variance was run to determine the effect of virtual audience interactivity on the user's perception of the training application. Four measures of user's perception were assessed: feeling of social presence, fear, confidence as a speaker, and perceived realism. A multivariate analysis of variance revealed that there was a statistically significant difference between interactive and non-interactive virtual audiences on the combined dependent variables, $F(4, 52) = 3.25$, $p = .019$; Wilks' $\Lambda = .800$; partial $\eta^2 = .200$.

The realism of the VR training application was rated high for both interactive and non-interactive conditions, which lead to a non-significant difference between the groups. Feeling fear in interactive conditions was evaluated higher than in non-interactive conditions. However, the result was not statistically significant. This is explained by the fact that the people with a pronounced fear of public speaking were not recruited (further research should take such cases into account).

The confidence of the speaker was lower during a performance in front of the interactive virtual audience than during performance in front of the non-interactive virtual audience, but without statistical significance as well. It is explained by the fact

that most of the participants were aware of the environment (students of the university and staff of the university who took part in the study were presented in the same classroom regularly). Further research should examine participants who are not familiar with the environment or do not make presentations regularly.

Finally, the feeling of presence was significantly higher for the interactive condition rather than for non-interactive conditions. It is supporting the statement that the interactivity of the VR application changes users' perception. Interactivity should be taken into account and be improved in VR public speaking training applications to increase the feeling of presence and during the training and, as a result, increase the effectiveness and success of the training. These findings are essential for the state of research and can be used by researchers in further improvements in VR training applications.

5.2. Limitations and Strengths

One general limitation of the study is that the topic of public speaking presentations was always limited to one subject as a description of some city. All observation, implementation, and evaluation studies always included presentations and questions regarding the same topic. However, this research can be a good example for further implementations (e.g., for including the wider variety of topics into the training applications).

The second limitation is that the study was focused on the fear of public speaking and did not consider the anxiety cases. Therefore, it remains an open question of how phobic people behave in front of the audience and how phobic participants react to the VR public speaking application. Anxiety cases should be taken into account

in future studies to improve the application and make it efficient for a broader range of prospective users.

Both observation and evaluation studies were conducted with participants who had an academic background (students and university staff members). Further research should involve different target groups (e.g., politicians, purples). Different target groups might behave differently in public sparking situations as well as perceive VR public speaking applications differently.

Finally, some of the technologies used in the research were improved during the study, or even newer technologies become available. Future work of this kind may use even more advanced technology, for example, 3D-records or live 3D-streams of real persons produced by 3D-depth cameras (Beck & Froehlich, 2017). Another possible further step of research can be the use of improved recognition tools.

Despite these limitations, the study has several strengths. The unique interactive virtual audience was created: the VR audience mostly behaved autonomously, in ways that were generated spontaneously and not regulated by an online operator or prepared in advance. The behavior of the interactive VR audience depended on the speaker's behavior, except for the verbal interactions. Although an operator still regulated the questions, all the interactive behavior patterns were the outcomes of a real audience behavior observation.

The knowledge obtained in the present dissertation thesis has several practical implications. First, a detailed understanding of how real audience members and real speakers interact with each other during public speaking situations was introduced. Second, the efficient ways to implement the identified interactive behavior patterns

between real audience and real speaker into the VR public speaking application were found and described. Finally, the new interactive virtual audiences created by the use of advanced technical equipment and with interactive behavior lead to a higher level of feeling of presence. The study contributes to a better understanding of participants' perceptions of interactive VR public speaking training application. In practice, it is recommended for developers to take interactivity into account when trying to achieve a higher feeling of presence during the VR trainings.

References

Abner, N., Cooperrider, K., & Goldin-Meadow, S. (2015). Gesture for Linguists: A Handy Primer. *Language and Linguistics Compass*, *9*(11), 437–451. https://doi.org/10.1111/lnc3.12168

Adler, R. B., & Towne, N. (1996). *Looking outlooking in: Interpersonal communication* (8th ed.). Fort Worth: Harcourt Brace College Publishers.

Affectiva. (2016). Emotion Behind Facial Expressions. Retrieved from http://blog.affectiva.com/the-emotion-behind-facial-expressions

Aldrich, C. (2004). *Simulations and the future of learning: An innovative (and perhaps revolutionary) approach to e-learning / Clark Aldrich*. San Francisco, Calif., Great Britain: Pfeiffer. Retrieved from http://www.loc.gov/catdir/bios/wiley046/2003009008.html

Ali, A., & Aggarwal, J. K. (2001). Segmentation and recognition of continuous human activity. In *Proceedings IEEE Workshop on Detection and Recognition of Events in Video* (pp. 28–35). IEEE Comput. Soc. https://doi.org/10.1109/EVENT.2001.938863

Allbeck, J. M., & Badler, N. I. (1998). Avatars a` la Snow Crash. In *Proceedings Computer Animation '98 (Cat. No.98EX169)* (pp. 19–24). IEEE Comput. Soc. https://doi.org/10.1109/CA.1998.681903

Allwood, J., & Cerrato, L. (2003). A study of gestural feedback expressions.

Altman, D. G. (1990). *Practical statistics for medical research.*

American National Standards Institute. (1973). *American national psychoacoustical terminology*. New York: Anonymous.

Ananthakrishnan, S., & Narayanan, S. (2009). Unsupervised Adaptation of Categorical Prosody Models for Prosody Labeling and Speech Recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, *17*(1), 138–149. https://doi.org/10.1109/TASL.2008.2005347

Anderson, P., Rothbaum, B. O., & Hodges, L. F. (2003). Virtual reality exposure in the treatment of social anxiety. *Cognitive and Behavioral Practice*, *10*(3), 240–247. https://doi.org/10.1016/S1077-7229(03)80036-6

Anderson, P. L., Zimand, E., Hodges, L. F., & Rothbaum, B. O. (2005). Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and Anxiety*, *22*(3), 156–158. https://doi.org/10.1002/da.20090

Aukstakalnis, S., & Blatner, D. (1992). *Silicon mirage: The art and science of virtual reality / by Steve Aukstakalnis and David Blatner.*

Baieier, K.P. (1993). Virtual reality: Short introduction.: [Online].

Bailenson, J. N., & Yee, N. (2006). A Longitudinal Study of Task Performance, Head Movements, Subjective Report, Simulator Sickness, and Transformed Social Interaction in Collaborative Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *15*(6), 699–716. https://doi.org/10.1162/pres.15.6.699

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction*. Cambridge: Cambridge University Press.

Bakhtin, M. (1999). The problem of speech genres. In A. Jaworski & N. Coupland (Eds.), *The discourse reader* (2nd ed., pp. 60–102). London, New York: Routledge.

Barfield, W., & Weghorst, S. (2009). The Sense of Presence within Virtual Environments: A Conceptual Framework. In M. J. Smith & G. Salvendy (Eds.),

Lecture notes in computer science, 0302-9743: 5617-5618. Human interface and the management of information: Symposium on Human Interface 2009 held as part of HCI International 2009, San Diego, CA, USA, July 19-24 2009, proceedings / Michael J. Smith, Gavriel Salvendy, eds (pp. 699–704). Berlin: Springer.

Baringer, D. K., & McCroskey, J. C. (2000). Immediacy in the classroom: Student immediacy. *Communication Education*, *49*(2), 178–186. https://doi.org/10.1080/03634520009379204

Bartholomay, E. M., & Houlihan, D. D. (2016). Public Speaking Anxiety Scale: Preliminary psychometric data and scale validation. *Personality and Individual Differences*, *94*, 211–215. https://doi.org/10.1016/j.paid.2016.01.026

Batrinca, L., Stratou, G., Shapiro, A., Morency, L.-P., & Scherer, S. (2013). Cicero - Towards a Multimodal Virtual Audience Platform for Public Speaking Training. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell,. . . H. Shimodaira (Eds.), *Lecture Notes in Computer Science. Intelligent Virtual Agents* (Vol. 8108, pp. 116–128). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40415-3_10

Beattie, G., & Shovelton, H. (2016). Mapping the Range of Information Contained in the Iconic Hand Gestures that Accompany Spontaneous Speech. *Journal of Language and Social Psychology*, *18*(4), 438–462. https://doi.org/10.1177/0261927X99018004005

Beck, A. T., Emery, G., & Greenberg, R. L. (2005). *Anxiety disorders and phobias: A cognitive perspective / Aaron T. Beck and Gary Emery with Ruth L. Greenberg ;*

[with a new preface by Aaron T. Beck] ([15th anniversary ed.], rev. pbk. ed.). Cambridge, Mass.: Basic Books.

Beck, S., & Froehlich, B. (2017). Sweeping-based volumetric calibration and registration of multiple RGBD-sensors for 3D capturing systems. In *2017 IEEE Virtual Reality (VR): Proceedings: March 18-22, 2017, Los Angeles, CA, USA* (pp. 167–176). Piscataway, NJ: IEEE. https://doi.org/10.1109/VR.2017.7892244

Berelson, B. (1984). *Content analysis in communication research* (Facsimile). New York: Hafner Press.

Berlo, D. K. (1974). *The process of communication: An introduction to theory and practice* (17th pr). New York, N.Y., [etc.]: Holt, Rinehart, and Winston.

Beyond Verbal. (2017): Beyond Verbal Developers. Retrieved from http://developers.beyondverbal.com/Account/Login?ReturnUrl=%2FHome%2Fapi

Biocca, F., Harms, C., & Gregg, J. (2001). The networked mind measure of social presence: Pilot test of the factor structure and concurrent validity. *4th Annual International Work-Shop on Presence, Philadelphia*, 1–9.

Biocca, F. (1997). The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments [1]. *Journal of Computer-Mediated Communication*, *3*(2), 0. https://doi.org/10.1111/j.1083-6101.1997.tb00070.x

Bishop, G., & Fuchs, H. (1992). Research directions in virtual environments. *ACM SIGGRAPH Computer Graphics*, *26*(3), 153–177. https://doi.org/10.1145/142413.142416

Bissonnette, J., Dubé, F., Provencher, M. D., & Moreno Sala, M. T. (2015). Virtual Reality Exposure Training for Musicians: Its Effect on Performance Anxiety and

Quality. *Medical Problems of Performing Artists*, *30*(3), 169–177. https://doi.org/10.21091/mppa.2015.3032

Biswas, K. K., & Basu, S. K. (2011). Gesture recognition using Microsoft Kinect®. In *The 5th International Conference on Automation, Robotics, and Applications* (pp. 100–103). IEEE. https://doi.org/10.1109/ICARA.2011.6144864

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). TARGET ARTICLE: Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychological Inquiry*, *13*(2), 103–124. https://doi.org/10.1207/S15327965PLI1302_01

Bodie, G. D. (2010). A Racing Heart, Rattling Knees, and Ruminative Thoughts: Defining, Explaining, and Treating Public Speaking Anxiety. *Communication Education*, *59*(1), 70–105. https://doi.org/10.1080/03634520903443849

Botella, C., Baños, R. M., Perpiñá, C., Villa, H., Alcañiz, M., & Rey, A. (1998). Virtual reality treatment of claustrophobia: A case report. *Behaviour Research and Therapy*, *36*(2), 239–246.

Bouchard, S., Dumoulin, S., Robillard, G., Guitard, T., Klinger, É., Forget, H., . . . Roucaut, F. X. (2017). Virtual reality compared with in vivo exposure in the treatment of social anxiety disorder: A three-arm randomised controlled trial. *The British Journal of Psychiatry: the Journal of Mental Science*, *210*(4), 276–283. https://doi.org/10.1192/bjp.bp.116.184234

Bowman, D. A. (1998). Interaction techniques for immersive virtual environments: Design, evaluation, and application. *Journal of Visual Languages and Computing*, *10*, 37–53.

Bowman, D. A., & McMahan, R. P. (2007b). Virtual Reality: How Much Immersion Is Enough? *Computer*, *40*(7), 36–43. https://doi.org/10.1109/MC.2007.257

Bowyer, K. W., Chang, K., & Flynn, P. (2006). A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. *Computer Vision and Image Understanding*, *101*(1), 1–15. https://doi.org/10.1016/j.cviu.2005.05.005

Brandenburg, K., Brix, S., & Sporer, T. (Eds.) 2004. *Wave Field Synthesis: From Research to Applications.*

Bray, J., & Maxwell, S. (1985). *Multivariate Analysis of Variance*. 2455 Teller Road, Newbury Park California 91320 United States of America: SAGE Publications, Inc.

Broaders, S. C., & Goldin-Meadow, S. (2010). Truth is at hand: How gesture adds information during investigative interviews. *Psychological Science*, *21*(5), 623–628. https://doi.org/10.1177/0956797610366082

Brooks, F. P. (1999). What's real about virtual reality? *IEEE Computer Graphics and Applications*, *19*(6), 16–27. https://doi.org/10.1109/38.799723

Burdea, G., & Coiffet, P. (2017). *Virtual reality technology [electronic resource]* (2nd ed.). Indianapolis, Indiana: [Wiley-IEEE Press].

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

Cavazza, M., Lugrin, J.-L., Hartley, S., Le Renard, M., Nandi, A., Jacobson, J., & Crooks, S. (2005). Intelligent virtual environments for virtual reality art. *Computers & Graphics*, *29*(6), 852–861. https://doi.org/10.1016/j.cag.2005.09.002

Chapanis, A. (1991). To Communicate the Human Factors Message, You Have to Know What the Message Is and How to Communicate It. *Human Factors Society Bulletin*, *34*(11), 1–4.

Chaplin, W. F., John, O. P., & Goldberg, L. R. (1988). Conceptions of states and traits: Dimensional attributes with ideals as prototypes. *Journal of Personality and Social Psychology*, *54*(4), 541–557. https://doi.org/10.1037/0022-3514.54.4.541

Chavan, U. B., & Kulkarni, D. B. (2013). Facial Expression Recognition- Review. *International Journal of Latest Trends in Engineering and Technology (IJLTET)*, *3*(1), 237–243.

Chibelushi, C.C., & Bourel, F. (2003). Facial expression recognition: A brief tutorial overview. *CVonline: on-Line Compendium of Computer Vision*, *9*.

Chollet, M., Sratou, G., Shapiro, A., Morency, L., & Scherer, S. An interactive virtual audience platform for public speaking training. In *2014 International Conference on Autonomous Agents and Multi-agent Systems* (pp. 1657–1658). Retrieved from https://www.researchgate.net/publication/287164645_An_interactive_virtual_audience_platform_for_public_speaking_training

Chollet, M., Ochs, M., & Pelachaud, C. (2014). From Non-verbal Signals Sequence Mining to Bayesian Networks for Interpersonal Attitudes Expression. In T. Bickmore, S. Marsella, & C. Sidner (Eds.), *Lecture Notes in Computer Science. Intelligent Virtual Agents* (Vol. 8637, pp. 120–133). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-09767-1_15

Chollet, M., Wörtwein, T., Morency, L.-P., Shapiro, A., & Scherer, S. (2015). Exploring feedback strategies to improve public speaking. In K. Mase, M. Langheinrich, D.

Gatica-Perez, H. Gellersen, T. Choudhury, & K. Yatani (Eds.), *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15* (pp. 1143–1154). New York, New York, USA: ACM Press. https://doi.org/10.1145/2750858.2806060

Cochran, W. G. (1954). Some Methods for Strengthening the Common χ 2 Tests. *Biometrics*, *10*(4), 417. https://doi.org/10.2307/3001616

Cohn, J., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the facial action coding system. In J. A. Coan & J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment.* New York: Oxford University Press.

Curtis, K., Jones, G. J.F., & Campbell, N. (2015). Effects of Good Speaking Techniques on Audience Engagement. In Z. Zhang, P. Cohen, D. Bohus, R. Horaud, & H. Meng (Eds.), *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (pp. 35–42). New York, NY: ACM. https://doi.org/10.1145/2818346.2820766

Dael, N., Mortillaro, M., & Scherer, K. R. (2012). The Body Action and Posture Coding System (BAP): Development and Reliability. *Environmental Psychology and Nonverbal Behavior*, *36*(2), 97–121. https://doi.org/10.1007/s10919-012-0130-0

Daly, J. A., Vangelisti, A. L., & Lawrence, S. G. (1989). Self-focused attention and public speaking anxiety. *Personality and Individual Differences*, *10*(8), 903–913. https://doi.org/10.1016/0191-8869(89)90025-1

Dani, T.H., & Rajit, G. (1998). Virtual Reality - A New Technology for the Mechanical Engineer. In M. Kutz (Ed.), *Mechanical engineers' handbook* (2nd ed., pp. 319–327). New York, Chichester: Wiley.

Das, S. (2012). Speech Recognition Technique: A Review. *International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622*, *2*(3), 2071–2087.

Davis, K. H., Biddulph, R., & Balashek, S. (1952). Automatic Recognition of Spoken Digits. *The Journal of the Acoustical Society of America*, *24*(6), 637–642. https://doi.org/10.1121/1.1906946

De Greef, P., & IJsselsteijn, W. (2000). Social Presence in the PhotoShare Tele-Application. *PRESENCE 2000 - 3rd International Workshop on Presence, Delft, The Netherlands*.

Dellaert, F., Polzin, T., & Waibel, A. (1996). Recognizing emotion in speech. In H. T. E. Bunnell & W. Idsardi (Eds.), *ICSLP 96: Proceedings, fourth international conference on spoken language processing / sponsored by University of Delaware, Alfred I. duPont Institute, in conjunction with Acoustical Society of America ... [et al.]* (pp. 1970–1973). New York: Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ICSLP.1996.608022

Delmonte, M. M., & Ryan, G. M. (1983). The Cognitive-Somatic Anxiety Questionnaire (CSAQ): A factor analysis. *British Journal of Clinical Psychology*, *22*(3), 209–212. https://doi.org/10.1111/j.2044-8260.1983.tb00601.x

DeVito, J. A. (1986). *The communication handbook: A dictionary / Joseph A. Devito*. New York, London: Harper & Row.

Dickson, D. C. M. (2005). Preface. In D. C. M. Dickson (Ed.), *Insurance Risk and Ruin* (pp. xi–xii). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511624155.001

Doering, N., Bortz, J., Poeschl, S., Werner, C. S., Schermelleh-Engel, K., Gerhard, C., & Gäde, J. C. (2015). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (Fifth edition). *Springer-Lehrbuch*. Berlin, Heidelberg: Springer.

Dubiago, M., Poeschl, S., & Doering, N. (2017). On the Effectiveness of a Virtual Public Speaking Influence of simulation fidelity on perceived realism. An exploratory study in a virtual public speaking training application. *The 22. CyberPsychology, CyberTherapy, and Social Networking Konferenz (CYPSY22), Wolverhampton, UK*.

Durlach, N. I., & Mavor, A. S. (1995). *Virtual reality: Scientific and technological challenges / Nathaniel I. Durlach and Anne S. Mavor, editors*. Washington, D.C.: National Academy Press.

Earnshaw, R. A., Gigante, M. A., & Jones, H. (1993). *Virtual reality systems*. London: Academic.

Efron, D. (1972). *Gesture, race, and culture: A tentative study of the Spatio-temporal and "linguistic" aspects of*. [Place of publication not identified]: Mouton De Gruyter.

Eisler, R. M., Miller, P. M., & Hersen, M. (1973). Components of assertive behavior. *Journal of Clinical Psychology*, *29*(3), 295–299.

Ekman, P. (1977). *Facial Expressions In Non-Verbal Behavior and Communication*: University of California, San Francisco.

Ekman, P., Friesen, W. V., & Hager, J. C. (1978). Facial Action Coding System (FACS). A Technique for the Measurement of Facial Action. *Consulting, Palo Alto,* (22).

Ekman, P., & Friesen, W. V. (1969). The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding. *Semiotica*, *1*(1). https://doi.org/10.1515/semi.1969.1.1.49

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129. https://doi.org/10.1037/h0030377

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, *1*(1), 56–75. https://doi.org/10.1007/BF01115465

Ekman, P., Friesen, W. V., & Tomkins, S. S. (1971). Facial Affect Scoring Technique: A First Validity Study. *Semiotica*, *3*(1). https://doi.org/10.1515/semi.1971.3.1.37

Elliott, R., Glauert, J. R. W., & Kennaway, J. R. (Eds.) 2004. *A Framework for Non-Manual Gestures in a Synthetic Signing System.*

Emmelkamp, P. (2013). Behavior Therapy with Adults. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change (6th ed.) // Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed.). Hoboken, N.J.: John Wiley & Sons.

Emmelkamp, P. M., Bruynzeel, M., Drost, L., & van der Mast, C. A. (2001). Virtual reality treatment in acrophobia: A comparison with exposure in vivo. *Cyberpsychology & Behavior: the Impact of the Internet, Multimedia, and Virtual*

Reality on Behavior and Society, *4*(3), 335–339. https://doi.org/10.1089/109493101300210222

Engel, R. J., & Schutt, R. K. (2014). *Fundamentals of Social Work Research* (Second Edition). Los Angeles: SAGE.

Fasel, B., & Luettin, J. (2003). Automatic facial expression analysis: a survey. *Pattern Recognition, 36*, 259–275.

FASP. (2014). Kompetenzzentrum Virtual Reality Ilmenau. Retrieved from.

Ferry, B., Kervin, L., Cambourne, B., Turbill, J., Puglisi, S., Jonassen, D., & Hedberg, J. (2004). Online classroom simulation: The 'next wave' for pre-service teacher education? *R. Atkinson, C. McBeath, D. Jonas-Dwyer & R. Phillips (Eds), Beyond the Comfort Zone: Proceedings of the 21st ASCILITE Conference*, *5*(8), 289–302. Retrieved from https://www.ascilite.org/conferences/perth04/procs/ferry.html

Feyereisen, P., Van de Wiele, M., & Dubois, F. (1988). The meaning of gestures: What can be understood without speech?, *8(1)*, 3–25.

Field, A. P. (2009). *Discovering statistics using SPSS: (and sex and drugs and rock 'n' roll) / Andy Field* (3rd ed.). *Introducing statistical methods*. Los Angeles i.e. Thousand Oaks, Calif., London: Sage Publications.

Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London: SAGE.

Fielding, M. L. (2006). *Effective communication in organisations* (3rd ed.). Lansdowne, Cape Town: Juta Academic.

Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin*, *99*(1), 20–35. https://doi.org/10.1037/0033-2909.99.1.20

Foley, J. D. (1994). *Introduction to computer graphics* (Abridged and modified version). Reading, Mass., Wokingham: Addison-Wesley.

Fonagy, I., & Magdics, K. (1963). Emotional Patterns in Intonation and Music. *STUF - Language Typology and Universals*, *16*(1-4). https://doi.org/10.1524/stuf.1963.16.14.293

Fraser, M., Glover, T., Vaghi, I., Benford, S., Greenhalgh, C., Hindmarsh, J., & Heath, C. (2000). Revealing the realities of collaborative virtual reality. In E. Churchill & M. Reddy (Eds.), *Proceedings of the third international conference on Collaborative virtual environments - CVE '00* (pp. 29–37). New York, New York, USA: ACM Press. https://doi.org/10.1145/351006.351010

Furmark, T., Tillfors, M., Stattin, H., Ekselius, L., & Fredrikson, M. (2000). Social phobia subtypes in the general population revealed by cluster analysis. *Psychological Medicine*, *30*(6), 1335–1344.

Furui, S. (2003). Recent advances in spontaneous Recent advances in spontaneous speech recognition and understanding. *Proc. IEEE Workshop on Spontaneous Speech Processing and Recognition*, 1–6.

Furui, S. (2005). Spontaneous speech recognition and summarization. *The Second Baltic Conference on HUMAN LANGUAGE*, 39–50. Retrieved from https://www.researchgate.net/publication/228348581_Spontaneous_speech_recognition_and_summarization

Gaikwad, S. K., Gawali, B. W., & Yannawar, P. (2010). A Review on Speech Recognition Technique. *International Journal of Computer Applications*, *10*(3), 16–24. https://doi.org/10.5120/1462-1976

Gandy, M., Starner, T., Auxier, J., & Ashbrook, D. (2000). The Gesture Pendant: A Self Illuminating, Wearable, Infrared Computer Vision System for Home Automation Control and Medical Monitoring. *IEEE Int. Symposium Wearable Computers*, 87–94.

Garcia-Garcia, J. M., Penichet, V. M. R., & Lozano, M. D. (2017). Emotion detection. In J. M. Gonzalez-Calleros (Ed.), *Proceedings of the XVIII International Conference on Human-Computer Interaction* (pp. 1–8). New York, NY: ACM. https://doi.org/10.1145/3123818.3123852

Gibbon, D., Mertins, I., & Moore, R. K. (2000). *Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology, and Product Evaluation*. *The Springer International Series in Engineering and Computer Science: Vol. 565*. Boston, MA: Springer US.

Gilkinson, H. (1942). Social fears as reported by students in college speech classes∗. *Speech Monographs*, *9*(1), 141–160. https://doi.org/10.1080/03637754209390068

Goetz, J., Kiesler, S., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003* (pp. 55–60). IEEE. https://doi.org/10.1109/ROMAN.2003.1251796

Goffman, E. (1963). *Behavior in public places*. [Place of publication not identified]: Free Press.

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, *64*, 257–283. https://doi.org/10.1146/annurev-psych-113011-143802

Goldin-Meadow, S., & Sandhofer, C. M. (1999). Gestures convey substantive information about a child's thoughts to ordinary listeners. *Developmental Science*, *2*(1), 67–74. https://doi.org/10.1111/1467-7687.00056

Good Vibrations. (2017): Good Vibrations Company B.V. –Recognize emotions directly from the voice. Retrieved from http://good-vibrations.nl/

Gratch, J., Rickel, J., Andre, E., Cassell, J., Petajan, E., & Badler, N. (2002). Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems*, *17*(4), 54–63. https://doi.org/10.1109/MIS.2002.1024753

Grewal, R., Cote, J. A., & Baumgartner, H. (2004). Multicollinearity and Measurement Error in Structural Equation Models: Implications for Theory Testing. *Marketing Science*, *23*(4), 519–529. https://doi.org/10.1287/mksc.1040.0070

Grother, P., Micheals, R. J., & Phillips, P. J. (2003). Face Recognition Vendor Test 2002 Performance Metrics. In G. Goos, J. Hartmanis, J. van Leeuwen, J. Kittler, & M. S. Nixon (Eds.), *Lecture Notes in Computer Science. Audio- and Video-Based Biometric Person Authentication* (Vol. 2688, pp. 937–945). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-44887-X_109

Gueguen, N., Jacob, C., & Martin, A. (2009). Mimicry in social interaction: Its effect on human judgment and behavior. *European Journal of Social Sciences*. (8 (2)), 253–259.

Gulzar, T., Singh, A., Rajoriya, D. K., & Farooq, N. (2014). A Systematic Analysis of Automatic Speech Recognition: An Overview. *Nternational Journal of Current Engineering and Technology*, *4*(3), 1664–1675.

Hamilton, C. (2011). *Communicating for results: A guide for business and the professions* (9th ed.). Boston MA: Wadsworth Cengage Learning.

Hargie, O. (2006). *The handbook of communication skills* (3rd ed.). London: Routledge.

Hargie, O., Dickson, D., & Tourish, D. (1999). *Communication in management*. Aldershot: Gower.

Harrigan, J., Rosenthal, R., & Scherer, K. (2008). *The New Handbook of Methods in Nonverbal Behavior Research*: Oxford University Press.

Harris, S. R., Kemmerling, R. L., & North, M. M. (2002). Brief virtual reality therapy for public speaking anxiety. *Cyberpsychology & Behavior: the Impact of the Internet, Multimedia, and Virtual Reality on Behavior and Society*, *5*(6), 543–550. https://doi.org/10.1089/109493102321018187

Hart, J., Gratch, J., & Marsella, S. (2017). How Virtual Reality Training Can Win Friends and Influence People. In G. Galanis, R. Sottilare, C. Best, & J. Kerry (Eds.), *Fundamental Issues in Defense Training and Simulation* (pp. 235–249). CRC Press. https://doi.org/10.1201/9781315583655-21

Heeter, C. (1992). Being There: The Subjective Experience of Presence. *Presence: Teleoperators and Virtual Environments*, *1*(2), 262–271. https://doi.org/10.1162/pres.1992.1.2.262

Heimberg, R. G., & Becker, R. E. (2002). *Cognitive-behavioral group therapy for social phobia: Basic mechanisms and clinical strategies / Richard G. Heimberg and Robert E. Becker*. New York, London: Guilford.

Hendrix, C., & Barfield, W. (1996). Presence within Virtual Environments as a Function of Visual Display Parameters. *Presence: Teleoperators and Virtual Environments, 5*(3), 274–289. https://doi.org/10.1162/pres.1996.5.3.274

Herrlich, M., Malaka, R., & Masuch, M. (Eds.). (2012). *Entertainment computing -- ICEC 2012: 11th International Conference, ICEC 2012, Bremen, Germany, September 26-29, 2012: proceedings / Marc Herrlich, Rainer Malaka, Maic Masuch (eds.). LNCS sublibrary. SL 3, Information systems and application, incl. Internet/Web and HCI: Vol. 7522*. Heidelberg: Springer.

Heylen, D. K. J. (2009). Understanding Speaker-Listener Interaction.

Heyman, R. E., Lorber, M. F., Mark Eddy, J., & West, T. V. (2014). Behavioral Observation and Coding. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 1–72). New York, NY: Cambridge University Press. Retrieved from http://psych.nyu.edu/westlab/documents/West_Chapter_2014_Behavioral%20observation%20and%20coding.pdf

Hills, T.W. (1993). In S. Bredekamp & T. J. Rosegrant (Eds.), *NAEYC: #225. Reaching potentials: Appropriate curriculum and assessment for young children / Sue Bredekamp and Teresa Rosegrant, editors. Volume 1.* Washington, DC: National Association for the Education of Young Children.

Hofmann, S. G., & Dibartolo, P. M. (2000). An instrument to assess self-statements during public speaking: Scale development and preliminary psychometric properties. *Behavior Therapy*, *31*(3), 499–515.

Hook, J. N., Smith, C. A., & Valentiner, D. P. (2008). A short-form of the Personal Report of Confidence as a Speaker. *Personality and Individual Differences*, *44*(6), 1306–1313. https://doi.org/10.1016/j.paid.2007.11.021

Hopkins, W. G. (2008). Research design: Choosing and fine-tuning a design for your study. *Sportscience*, *12*, 12–21.

Howell, D. C. (2010). *Statistical methods for psychology* (7th ed.). Australia, Belmont CA: Thomson Wadsworth.

Hudlicka, E. (2003). To feel or not to feel: The role of effect in human-computer interaction. *International Journal of Human-Computer Studies*, *59*(1-2), 1–32. https://doi.org/10.1016/S1071-5819(03)00047-8

Hummels, C., & Stappers, P. J. (1998). Meaningful gestures for human-computer interaction: Beyond hand postures. In *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 591–596). IEEE Comput. Soc. https://doi.org/10.1109/AFGR.1998.671012

Ijsselsteijn, W. A., Freeman, J., & Ridder, H. de. (2001). Presence: Where are we? *Cyberpsychology & Behavior: the Impact of the Internet, Multimedia, and Virtual Reality on Behavior and Society*, *4*(2), 179–182. https://doi.org/10.1089/109493101300117875

IJsselsteijn, W.A., Bouwhuis, D.G., & Ridder, d. H. (2004). *Presence in depth*. Eindhoven: Technische Universiteit Eindhoven.

International Society for Presence Research. (2000). The Concept of Presence: Explication Statement. Retrieved from https://ispr.info/about-presence-2/about-presence/

Isdale, J. (1998). What is virtual reality? A web-based introduction.

Johns-Lewis, C. (1986). *Intonation in discourse*. *Routledge library editions. Phonetics and phonology: volume 11*. Abingdon, Oxon, New York, NY: Routledge.

Jokinen, K., & Allwood, J. (2010). Hesitation in Intercultural Communication: Some Observations and Analyses on Interpreting Shoulder Shrugging. In T. Ishida (Ed.), *Lecture Notes in Computer Science. Culture and Computing* (Vol. 6259, pp. 55–70). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-17184-0_5

Juang, B. H. (1998). From speech recognition to understanding: Shifting the paradigm to achieve natural human-machine communication. *The Journal of the Acoustical Society of America*, *103*(5), 2817. https://doi.org/10.1121/1.421591

Kairos. (2017). Face Recognition, Emotion Analysis & Demographics. Retrieved from https://www.kairos.com/

Kang, N., Brinkman, W.-P., van Riemsdijk, M. B., & Neerincx, M. A. (2013). An Expressive Virtual Audiencewith Flexible Behavioral Styles. *IEEE Transactions on Affective Computing*, *4*(4), 326–340. https://doi.org/10.1109/TAFFC.2013.2297104

Kendon, A. (2002). Some uses of the head shake. *Gesture*, *2*(2), 147–182. https://doi.org/10.1075/gest.2.2.03ken

Khan, R. Z. (2012). Hand Gesture Recognition: A Literature Review. *International Journal of Artificial Intelligence & Applications*, *3*(4), 161–174. https://doi.org/10.5121/ijaia.2012.3412

Kleckova, J., & Mahdian, B. (2004). Nonverbal Communication in Spontaneous Speech Recognition. *WSEAS Transactions on Electronics*, *1*(3), 531–536.

Kothgassner, O.D., Felnhofer, A., Beutl, L., Hlavacs, H., Lehenbauer, M., & Stetina, B. (2012). A Virtual Training Tool for Giving Talks. In M. Herrlich, R. Malaka, & M. Masuch (Eds.), *LNCS sublibrary. SL 3, Information systems and application, incl. Internet/Web and HCI: Vol. 7522. Entertainment computing -- ICEC 2012: 11th International Conference, ICEC 2012, Bremen, Germany, September 26-29, 2012 : proceedings / Marc Herrlich, Rainer Malaka, Maic Masuch (eds.).* Heidelberg: Springer.

Kratz, L., Smith, M., & Lee, F. J. (2007). Wizards. In B. Kapralos, M. Katchabaw, & J. Rajnovich (Eds.), *Proceedings of the 2007 conference on Future Play - Future Play '07* (p. 209). New York, New York, USA: ACM Press. https://doi.org/10.1145/1328202.1328241

Krauss, R. M. (2002). Verbal Communication, Psychology of. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 16161–16165). Elsevier. https://doi.org/10.1016/B0-08-043076-7/01815-5

Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal Behavior and Nonverbal Communication: What do Conversational Hand Gestures Tell Us? In *Advances in Experimental Social Psychology. Advances in Experimental Social Psychology*

Volume 28 (Vol. 28, pp. 389–450). Elsevier. https://doi.org/10.1016/S0065-2601(08)60241-5

Kwon, J. H., Powell, J., & Chalmers, A. (2013). How the level of realism influence anxiety in virtual reality environments for a job interview. *International Journal of Human-Computer Studies*, *71*(10), 978–987. https://doi.org/10.1016/j.ijhcs.2013.07.003

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Larsson, P., Vastfjall, D., & Kleiner, M. (2002). Better presence and performance in virtual environments by improved binaural sound rendering. *Virtual, Synthetic, and Entertainment Audio*.

LaViola, J. J. (2017). *3D user interfaces: Theory and practice* (Second edition). Boston: Addison-Wesley.

Lecrubier, Y., Wittchen, H. U., Faravelli, C., Bobes, J., Patel, A., & Knapp, M. (2000). A European perspective on social anxiety disorder. *European Psychiatry*, *15*(1), 5–16. https://doi.org/10.1016/S0924-9338(00)00216-9

Lee, C., Rincon, G. A., Meyer, G., Höllerer, T., & Bowman, D. A. (2013). The effects of visual realism on search tasks in mixed reality simulation. *IEEE Transactions on Visualization and Computer Graphics*, *19*(4), 547–556. https://doi.org/10.1109/TVCG.2013.41

Lee, C. M., & Narayanan, S. S. (2005). Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing*, *13*(2), 293–303. https://doi.org/10.1109/TSA.2004.838534

Lee, C. M., Narayanan, S. S., & Pieraccini, R. (2002). Classifying emotions in human-machine spoken dialogs. In *2002 IEEE international conference on multimedia and expo: ICME 2002* (pp. 737–740). IEEE. https://doi.org/10.1109/ICME.2002.1035887

Lee, J. M., Ku, J. H., Jang, D. P., Kim, D. H., Choi, Y. H., Kim, I. Y., & Kim, S. I. (2002). Virtual reality system for treatment of the fear of public speaking using image-based rendering and moving pictures. *Cyberpsychology & Behavior: the Impact of the Internet, Multimedia, and Virtual Reality on Behavior and Society*, *5*(3), 191–195. https://doi.org/10.1089/109493102760147169

Lee, K.-M., & Nass, C. (2005). Social-Psychological Origins of Feelings of Presence: Creating Social Presence With Machine-Generated Voices. *Media Psychology*, *7*(1), 31–45. https://doi.org/10.1207/S1532785XMEP0701_2

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators and Virtual Environments*, *10*(3), 282–297. https://doi.org/10.1162/105474601300343612

Levis, J. M., & Suvorov, R. (2013). Automatic Speech Recognition. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics.* Chichester, West Sussex, UK: Wiley-Blackwell.

Levis, J. M. (2013). Suprasegmentals: Intonation. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 5443–5449). Chichester, West Sussex, UK: Wiley-Blackwell.

Lewis, C. H., & Griffin, M. J. (1997). Human factors consideration in clinical applications of virtual reality. *Studies in Health Technology and Informatics*, *44*, 35–56.

Li, B. Y.L., Mian, A. S., Liu, W., & Krishna, A. (2013). Using Kinect for face recognition under varying poses, expressions, illumination, and disguise. In *IEEE Workshop on Applications of Computer Vision (WACV), 2013: [17th - 18th January 2013], Clearwater Beach, FL, USA ; [part of the IEEE Winter Vision Meetings 2013 (WVM)], 15 - 17 [i.e. 16 - 18] Jan. 2013* (pp. 186–192). Piscataway, NJ: IEEE. https://doi.org/10.1109/WACV.2013.6475017

Ling, Y., Nefs, H. T., Morina, N., Heynderickx, I., & Brinkman, W.-P. (2014). A meta-analysis on the relationship between self-reported presence and anxiety in virtual reality exposure therapy for anxiety disorders. *PloS One*, *9*(5), e96144. https://doi.org/10.1371/journal.pone.0096144

Lombard, M., Ditton, T. B., Crane, D., Davis, B., Gil-Egui, G., Horvath, K., . . . Park, S. (2000). Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument. *Third International Workshop on Presence*. (Delft, The Netherlands.).

Lombard, M., & Ditton, T. (1997). At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, *3*(2), 0. https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94–100. https://doi.org/10.1016/j.chb.2014.04.043

Lucas, S. (2009). *The art of public speaking* (Tenth edition). Boston: McGraw-Hill Higher Education.

Ludlow, P. (1996). *High noon on the electronic frontier: Conceptual issues in cyberspace / Peter Ludlow*. *Digital communication*. Cambridge, Mass., London: MIT Press.

Lugovic, S., Dunder, I., & Horvat, M. (2016). Techniques and applications of emotion recognition in speech. In *2016 39th International Convention on Information and Communication Technology, Electronics, and Microelectronics (MIPRO)* (pp. 1278–1283). IEEE. https://doi.org/10.1109/MIPRO.2016.7522336

Ma, B., Xu, W., & Wang, S. (2013). A Robot Control System Based on Gesture Recognition Using Kinect. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, *11*(5). https://doi.org/10.11591/telkomnika.v11i5.2493

MacIntyre, P. D., Thivierge, K. A., & MacDonald, J. R. (1997). The effects of audience interest, responsiveness, and evaluation on public speaking anxiety and related variables. *Communication Research Reports*, *14*(2), 157–168. https://doi.org/10.1080/08824099709388657

Madabhushi, A., & Aggarwal, J. K. (2000). Using head movement to recognize the activity. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000* (pp. 698–701). IEEE Comput. Soc. https://doi.org/10.1109/ICPR.2000.903013

Mann, S., & Robinson, A. (2009). Boredom in the lecture theatre: An investigation into the contributors, moderators, and outcomes of boredom amongst university students. *British Educational Research Journal*, *35*(2), 243–258. https://doi.org/10.1080/01411920802042911

Marteau, T. M., & Bekker, H. (1992b). The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, *31*(3), 301–306. https://doi.org/10.1111/j.2044-8260.1992.tb00997.x

McClave, E. (1994). Gestural beats: The rhythm hypothesis. *Journal of Psycholinguistic Research*, *23*(1), 45–66. https://doi.org/10.1007/BF02143175

McCreery, M. P., Schrader, P. G., Krach, S. K., & Boone, R. (2013). A sense of self: The role of presence in virtual environments. *Computers in Human Behavior*, *29*(4), 1635–1640. https://doi.org/10.1016/j.chb.2013.02.002

McCroskey, J. C. (2006). *An introduction to rhetorical communication: A western cultural perspective* (9th ed.). Boston MA: Allyn and Bacon.

Mcfall, M. E., Winnett, R. L., Bordewick, M. C., & Bornstein, P. H. (2016). Nonverbal Components in the Communication of Assertiveness. *Behavior Modification*, *6*(1), 121–140. https://doi.org/10.1177/01454455820061008

McNeill, D. (1995). *Hand and mind: What gestures reveal about thought* (Pbk ed.). Chicago, London: University of Chicago.

McQuail, D. (1997). *Audience analysis*. Thousand Oaks, Calif.: Sage Publications.

Meehan, M., Insko, B., Whitton, M., & Brooks, F. P. (2002). Physiological measures of presence in stressful virtual environments. In T. Appolloni (Ed.), *ACM SIGGRAPH 2002* (p. 645). New York, New York, USA: ACM Press. https://doi.org/10.1145/566570.566630

Mehrabian, A. (1968). Relationship of attitude to seated posture, orientation, and distance. *Journal of Personality and Social Psychology*, *10*(1), 26–30. https://doi.org/10.1037/h0026384

Mehrabian, A. (1971). *Silent messages.*

Menzel, K. E., & Carrell, L. J. (1994). The relationship between preparation and performance in public speaking. *Communication Education*, *43*(1), 17–26. https://doi.org/10.1080/03634529409378958

Meyers, L. S., Gamst, G., & Guarino, A. J. (2013). *Applied multivariate research: Design and interpretation / Lawrence S. Meyers, Glenn Gamst, and A.J. Guarino* (2nd ed.). Thousand Oaks, Calif., London: SAGE.

Microsoft. (2017): Microso Cognitive Services – Emotion API. Retrieved from h ps://www.microso .com/cognitive-services/en-us/emotion-api

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook / Matthew B. Miles, A. Michael Huberman* (2nd ed.). Thousand Oaks, Calif., London: SAGE.

Morris, L. W., Davis, M. A., & Hutchings, C. H. (1981). Cognitive and emotional components of anxiety: Literature review and a revised worry-emotionality scale. *Journal of Educational Psychology*, *73*(4), 541–555.

Myers, S. A., & Ferry, M. F. (2001). Interpersonal communication motives and nonverbal immediacy behaviors. *Communication Research Reports*, *18*(2), 182–191. https://doi.org/10.1080/08824090109384796

Naumovski, V., Dana, L.-P., Pesakovic, G., & Fidanoski, F. (2017). Why interpersonal communication is important in public administration? *Współczesne Problemy Ekonomiczne, 14*, 55–77. https://doi.org/10.18276/wpe.2017.14-04

Navarro, J., & Karlins, M. (2008, 2007). *What every BODY is saying [electronic resource]: An ex-FBI agent's guide to speed reading people / by Joe Navarro, with Marvin Karlins*. New York, NY: HarperCollins.

Neiva, D. H., & Zanchettin, C. (2016). A Dynamic Gesture Recognition System to Translate between Sign Languages in Complex Backgrounds. In *2016 5th Brazilian Conference on Intelligent Systems (BRACIS)* (pp. 421–426). IEEE. https://doi.org/10.1109/BRACIS.2016.082

Nolan, F. (2006). Intonation. In B. Aarts & A. McMahon (Eds.), *The Handbook of English Linguistics* (pp. 433–457). Malden, MA, USA: Blackwell Publishing. https://doi.org/10.1002/9780470753002.ch19

North, M. M., North, S. M., & Coble, J. R. (1997). Virtual reality therapy: An effective treatment for psychological disorders. *Studies in Health Technology and Informatics, 44*, 59–70.

NViso. (2016). Artificial Intelligence Emotion Recognition Software. http://nviso.ch/. Retrieved from 18.03.2019

Oh, S. Y., Bailenson, J., Krämer, N., & Li, B. (2016). Let the Avatar Brighten Your Smile: Effects of Enhancing Facial Expressions in Virtual Environments. *PloS One, 11*(9), e0161794. https://doi.org/10.1371/journal.pone.0161794

Okechukwu, M., & Udoka, F. (2011). Understanding Virtual Reality Technology: Advances and Applications. In M. Schmidt (Ed.), *Advances in Computer Science and Engineering.* InTech. https://doi.org/10.5772/15529

Oland, A., & Dannenberg, R. (2017). Loudness Concepts & Pan Laws: Introduction to Computer Music Carnegie Mellon University. Retrieved from http://www.cs.cmu.edu/~music/icm-online/readings/panlaws/panlaws.pdf

Paeschke, A., & Sendlmeier, W. F. (2000). Prosodic Characteristics of Emotional Speech: Measurements of Fundamental Frequency Movements. *Presented at Proceedings of the ISCA-Workshop on Speech and Emotion*.

Parsons, T. D., & Rizzo, A. A. (2008). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, *39*(3), 250–261. https://doi.org/10.1016/j.jbtep.2007.07.007

Paul, G. L. (1967). Insight versus desensitization in psychotherapy two years after termination. *Journal of Consulting Psychology*, *31*(4), 333–348. https://doi.org/10.1037/h0024855

Pelachaud, C. (2005). Multimodal expressive embodied conversational agents. In H. Zhang, T.-S. Chua, R. Steinmetz, M. Kankanhalli, & L. Wilcox (Eds.), *Proceedings of the 13th annual ACM international conference on Multimedia - MULTIMEDIA '05* (p. 683). New York, New York, USA: ACM Press. https://doi.org/10.1145/1101149.1101301

Pereira, F. C. N., & Ebrahimi, T. (2002). *The MPEG-4 book. IMSC Press multimedia series*. Upper Saddle River, NJ, Great Britain: Prentice Hall PTR.

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

Pertaub, D.-P., Slater, M., & Barker, C. (2002). An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience. *Presence: Teleoperators and Virtual Environments*, *11*(1), 68–78. https://doi.org/10.1162/105474602317343668

Picard, R. W. (1999). *Affective Computing for HCI.* HCI (1). Retrieved from file:///C:/Users/Maria/Downloads/cis_14_3975.pdf

Poeschl, S. (2017). Virtual Reality Training for Public Speaking—A QUEST-VR Framework Validation. *Frontiers in ICT*, *4*, 944. https://doi.org/10.3389/fict.2017.00013

Poeschl, S., & Doering, N. (2012a). Designing virtual audiences for fear of public speaking training - an observation study on realistic nonverbal behavior. *Studies in Health Technology and Informatics*, *181*, 218–222.

Poeschl, S., & Doering, N. (2012b). Virtual training for Fear of Public Speaking — Design of an audience for immersive virtual environments. In *2012 IEEE Virtual Reality (VR)* (pp. 101–102). IEEE. https://doi.org/10.1109/VR.2012.6180902

Poeschl, S., & Doering, N. (2013). The German VR Simulation Realism Scale – Psychometric Construction for Virtual Reality Applications with Virtual Humans. *Annual Review of Cybertherapy and Telemedicine 2013, Interactive Media Institute, and IOS Press.* Retrieved from https://www.researchgate.net/publication/241693096_The_German_VR_Simulation_

Realism_Scale_-_Psychometric_Construction_for_Virtual_Reality_Applications_with_Virtual_Humans

Poeschl, S., & Doering, N. (2015). Measuring Co-Presence and Social Presence in Virtual Environments - Psychometric Construction of a German Scale for a Fear of Public Speaking Scenario. *Annual Review of CyberTherapy and Telemedicine, 13*, 58–63. Retrieved from https://www.researchgate.net/publication/291517481_Measuring_Co-Presence_and_Social_Presence_in_Virtual_Environments_-_Psychometric_Construction_of_a_German_Scale_for_a_Fear_of_Public_Speaking_Scenario

Poeschl, S., Tudor, A.-D., & Doering, N. (2014). Human Factors in IVE Development - A Case Study For Virtual Fear of Public Speaking Training.

Porter, K., Porcari, C., Koch, E. I., Fons, C., & Spates, C. R. (2006). In vivo exposure treatment for agoraphobia. *The Behavior Analyst Today, 7*(3), 434–441. https://doi.org/10.1037/h0100161

Powers, M. B., & Emmelkamp, P. M. G. (2008). Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders, 22*(3), 561–569. https://doi.org/10.1016/j.janxdis.2007.04.006

Poyatos, F. (1985). The deeper levels of face-to-face interaction. *Language & Communication*, *5*(2), 111–131. https://doi.org/10.1016/0271-5309(85)90004-7

Pribyl, C. B., Keaten, J., & Sakamoto, M. (2001). The effectiveness of a skills-based program in reducing public speaking anxiety. *Japanese Psychological Research*, *43*(3), 148–155. https://doi.org/10.1111/1468-5884.t01-1-00171

Price, M., & Anderson, P. (2007). The role of presence in virtual reality exposure therapy. *Journal of Anxiety Disorders*, *21*(5), 742–751. https://doi.org/10.1016/j.janxdis.2006.11.002

Pujol-Tost, L. (2018). Cultural Presence in Virtual Archaeology: An Exploratory Analysis of Factors. *Presence: Teleoperators and Virtual Environments*, *26*(03), 247–263. https://doi.org/10.1162/PRES_a_00296

Qu, C. (2014). *Talking with a Virtual Human: Controlling the Human Experience and Behavior in a Virtual Conversation.*

Rabiner, L. R., & Juang, B. H. (1993). *Fundamentals of speech recognition*. *Prentice-Hall signal processing series*. Englewood Cliffs, NJ: PTR Prentice Hall; London: Prentice-Hall International (UK).

Ragan, E. D., Bowman, D. A., Kopper, R., Stinson, C., Scerbo, S., & McMahan, R. P. (2015). Effects of Field of View and Visual Complexity on Virtual Reality Training Effectiveness for a Visual Scanning Task. *IEEE Transactions on Visualization and Computer Graphics*, *21*(7), 794–807. https://doi.org/10.1109/TVCG.2015.2403312

Ragan, E. D., Wilkes, C., Cao, Y., & Bowman, D. A. (2012). The effects of virtual character animation on spatial judgments. In *2012 IEEE Virtual Reality (VR)* (pp. 141–142). IEEE. https://doi.org/10.1109/VR.2012.6180921

Rapee, R. M., & Heimberg, R. G. (1997). A cognitive-behavioral model of anxiety in social phobia. *Behaviour Research and Therapy*, *35*(8), 741–756. https://doi.org/10.1016/s0005-7967(97)00022-3

Remington, R. W., Johnston, J. C., & Yantis, S. (1992). Involuntary attentional capture by abrupt onsets. *Perception & Psychophysics*, *51*(3), 279–290. https://doi.org/10.3758/BF03212254

Ren, F. (2010). Emotion Recognition of Weblog Sentences Based on an Ensemble Algorithm of Multi-label Classification and Word Emotions, *2(1)*, 1–14.

Ricci Bitti, P. E., & Poggi, I. A. (1991). Symbolic nonverbal behavior: Talking through gestures. In R. S. Feldman & B. Rimé (Eds.), *Studies in emotion and social interaction. Fundamentals of nonverbal behavior.* Cambridge: Cambridge University Press.

Richmond, V. P. (2003). Teacher nonverbal immediacy: Use and outcomes. In J. L. Chesebro & J. McCroskey (Eds.), *Communication for teachers* (pp. 65–82). Boston: Allyn & Bacon.

Riffe, D., Lacy, S., & Fico, F. (2005). *Analyzing media messages: Using quantitative content analysis in research / Daniel Riffe, Stephen Lacy, Frederick G. Fico* (2nd ed.). *LEA's communication series*. Mahwah, N.J., Lawrence Erlbaum, London.

Riva, G., Alcañiz, M., Anolli, L., Bacchetta, M., Baños, R., Beltrame, F., . . . Troiañi, R. (2001). The VEPSY updated project: Virtual reality in clinical psychology. *Cyberpsychology & Behavior: the Impact of the Internet, Multimedia, and Virtual Reality on Behavior and Society*, *4*(4), 449–455. https://doi.org/10.1089/109493101750527006

Romano, J. M., & Bellack, A. S. (1980). Social validation of a component model of assertive behavior. *Journal of Consulting and Clinical Psychology*, *48*(4), 478–490.

Rose, Y. J., & Tryon, W. W. (2016). Judgments of Assertive Behavior as a Function of Speech Loudness, Latency, Content, Gestures, Inflection, and Sex. *Behavior Modification*, *3*(1), 112–123. https://doi.org/10.1177/014544557931008

Ross, R. S. (1989). *Speech communication: The speechmaking system / Raymond S. Ross* (8th ed.). Englewood Cliffs, N.J.: Prentice-Hall.

Rothbaum, B. O., Hodges, L. F., Kooper, R., Opdyke, D., Williford, J. S., & North, M. (1995). Effectiveness of computer-generated (virtual reality) graded exposure in the treatment of acrophobia. *The American Journal of Psychiatry*, *152*(4), 626–628. https://doi.org/10.1176/ajp.152.4.626

Rothbaum, B. O., Hodges, L., Smith, S., Lee, J. H., & Price, L. (2000). A controlled study of virtual reality exposure therapy for the fear of flying. *Journal of Consulting and Clinical Psychology*, *68*(6), 1020–1026.

Rui, Y., & Anandam, P. Segmenting visual actions based on spatiotemporal motion patterns: 2000. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 111–118). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.585.6627&rep=rep1&type=pdf

Safir, M. P., Wallach, H. S., & Bar-Zvi, M. (2012). Virtual reality cognitive-behavior therapy for public speaking anxiety: One-year follow-up. *Behavior Modification*, *36*(2), 235–246. https://doi.org/10.1177/0145445511429999

Samantaray, A., Nayak, S. K., & Mishra, A. K. (2013). Hand Gesture Recognition using Computer Vision. *International Journal of Scientific & Engineering Research, 4*(6), 1602–1609. Retrieved from https://pdfs.semanticscholar.org/e93b/51b7599365c540308e4271034a4802dd99d8.pdf

Sanders, R. (2003). Applying the skills concept to discourse and conversation: the remediation of performance defects in talk-in-interaction. In J. O. Greene & B. R. Burleson (Eds.), *LEA's communication series. Handbook of communication and social interaction skills.* Mahwah, N.J., London: L. Erlbaum Associates.

Schegloff, E. A. (1982). “Discourse as an interactional achievement: Some uses of "uh huh" and other things that come between sentences. In D. Tannen (Ed.), *0196-7207. Analyzing Discourse: Text and Talk: 32nd Annual Roundtable on Languages and Linguistics: Papers* (pp. 71–93).

Schlömer, T., Poppinga, B., Henze, N., & Boll, S. (2008). Gesture recognition with a Wii controller. In A. Schmidt, H. Gellersen, E. van den Hoven, A. Mazalek, P. Holleis, & N. Villar (Eds.), *Proceedings of the 2nd international conference on Tangible and embedded interaction - TEI '08* (p. 11). New York, New York, USA: ACM Press. https://doi.org/10.1145/1347390.1347395

Schroeder, R. (2002). Social Interaction in Virtual Environments: Key Issues, Common Themes, and a Framework for Research. In D. Diaper, C. Sanger, & R. Schroeder (Eds.), *Computer Supported Cooperative Work. The Social Life of Avatars* (pp. 1–18). London: Springer London. https://doi.org/10.1007/978-1-4471-0277-9_1

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments*, *10*(3), 266–281. https://doi.org/10.1162/105474601300343603

Shanbhogue, M., Kulkarni, S., Suprith, R., Tejas, K. I., & Nagarathna, N. (2016). A Study on Speech Recognition. *International Journal for Research in Applied Science & Engineering*, *4*(III).

Shanbhogue, M., Kulkarni, S., Suprith, R., Tejas, K., & Nagarathna, N. (2016). A Study on Speech Recognition. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, *4*(III).

Sharma, R., Neumann, U., & Kim, C. (2002). Emotion Recognition In Spontaneous Emotional Utterances From Movie Sequences.

Sharma, R., Zeller, M., Pavovic, V. I., Huang, T. S., Lo, Z., Chu, S., . . . Schulten, K. (1996). "Speech/Gesture Interface to a Visual Computing Environment for Molecular Biologists. *Proc. of ICPR'96*, *II*, 964–968.

Sheridan, T. B. (1992). Musings on Telepresence and Virtual Presence. *Presence: Teleoperators and Virtual Environments*, *1*(1), 120–126. https://doi.org/10.1162/pres.1992.1.1.120

Shibata, T. (2002). Head-mounted display. *Displays*, *23*(1-2), 57–64. https://doi.org/10.1016/S0141-9382(02)00010-0

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. London: Wiley.

Shrishrimal, P., Deshmukh, R., & Waghmare, B. V. (2012). Indian Language Speech Database: A Review. *International Journal of Computer Applications*, *47*(5), 17–21. https://doi.org/10.5120/7184-9893

Shrum, J. L., & Glisan, E. W. (2009). *Teacher's handbook: Contextualized language instruction* (4th ed.). Boston MA: Heinle Cengage.

Siegman, A. W., & Feldstein, S. (1987). *Nonverbal behavior and communication* (2nd ed.). Hillsdale, NJ: L. Erlbaum.

Slater, M., Pertaub, D.-P., & Steed, A. (1999). Public speaking in virtual reality: Facing an audience of avatars. *IEEE Computer Graphics and Applications*, *19*(2), 6–9. https://doi.org/10.1109/38.749116

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater, M., Pertaub, D.-P., Barker, C., & Clark, D. M. (2006). An experimental study on fear of public speaking using a virtual environment. *Cyberpsychology & Behavior : the Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, *9*(5), 627–633. https://doi.org/10.1089/cpb.2006.9.627

Slater, M., Sadagic, A., Usoh, M., & Schroeder, R. (2000). Small-Group Behavior in a Virtual and Real Environment: A Comparative Study. *Presence: Teleoperators and Virtual Environments*, *9*(1), 37–51. https://doi.org/10.1162/105474600566600

Slater, M., Usoh, M., & Steed, A. (1994). Depth of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *3*(2), 130–144. https://doi.org/10.1162/pres.1994.3.2.130

Slater, M., & Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *6*(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Slovák, P., & Fitzpatrick, G. (2015). Teaching and Developing Social and Emotional Skills with Technology. *ACM Transactions on Computer-Human Interaction*, *22*(4), 1–34. https://doi.org/10.1145/2744195

Smith, D. (2003). Five principles for research ethics, *Vol 34, No. 1*, 56.

Spence, S. H. (2003). Social skills training with children and adolescents. *Child and Adolescent Mental Health*, *8*, 84–96. Retrieved from http://www.cs.cmu.edu/~smrobert/indep_summer/cam.pdf

Spielberger, C. D. (1983). Manual for the State-Trait Anxiety Inventory (STAI). *Palo Alto, CA: Consulting Psychologists Press*.

Stankov, L., Kleitman, S., & Jackson, S. A. (2015). Measures of the Trait of Confidence. In *Measures of Personality and Social Psychological Constructs* (pp. 158–189). Elsevier. https://doi.org/10.1016/B978-0-12-386915-9.00007-3

Steed, A., Slater, M., Sadagic, A., & Tromp, J. (1999). Leadership and collaboration in virtual environments. *IEEE Virtual Reality*, 58–63.

Steed, A., & Schroeder, R. (2015). Collaboration in Immersive and Non-immersive Virtual Environments. In M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, & R. J.

Schaevitz (Eds.), *Immersed in Media* (pp. 263–282). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-10190-3_11

Stramler, J. H. (1993). *The dictionary for human factors, ergonomics*. Boca Raton, Fla., London: CRC.

Strangert, E. (1993). Speaking style and pausing. *PHONUM, Reports from the Department of Phonetics, University of Umea*.

Strickland, D. (1996). Virtual reality and mental disorders (panel). In J. Fujii (Ed.), *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques - SIGGRAPH '96* (pp. 509–510). New York, New York, USA: ACM Press. https://doi.org/10.1145/237170.237302

Strickland, D., Hodges, L., North, M., & Weghorst, S. (1997). Overcoming phobias by virtual exposure. *Communications of the ACM*, *40*(8), 34–39. https://doi.org/10.1145/257874.257881

Sverke, M. (2007). Quantitative investigations: Design, sampling, and analysis. In B. Gustavsson (Ed.), *The principles of knowledge creation: Research methods in the social // The principles of knowledge creation: Research methods in the social sciences / edited by Bengt Gustavsson* (pp. 23–45). Cheltenham: Edward Elgar.

Swerts, M., & Geluykens, R. (2016). Prosody as a Marker of Information Flow in Spoken Discourse. *Language and Speech*, *37*(1), 21–43. https://doi.org/10.1177/002383099403700102

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson Education.

Tian, Y., Kanade, T., & Cohn, J. F. (2001). Recognizing action units for facial expression analysis. *Pattern Analysis and Machine Intelligence, IEEE, 23*, 97-115.

Triplett, N. (1898). The Dynamogenic Factors in Pacemaking and Competition. *The American Journal of Psychology*, *9*(4), 507. https://doi.org/10.2307/1412188

Tromp, J., Bullock, A., Steed, A., Sadagic, A., Slater, M., & Frecon, E. (1998). Small group behaviour experiments in the Coven project. *IEEE Computer Graphics and Applications*, *18*(6), 53–63. https://doi.org/10.1109/38.734980

Tudor, A.-D., & Doering, N. (2016). *Development of a virtual audience concept for public speaking training in immersive virtual environments*. *Kommunikationswissenschaft interdisziplinär: Vol. 5*. Ilmenau, Ilmenau: Universitätsverlag Ilmenau.

Tudor, A.-D., Mustatea, I., Poeschl, S., & Doering, N. (2014). Responsive audiences — Nonverbal cues as reactions to a speaker's behavior. In *2014 IEEE Virtual Reality (VR)* (pp. 119–120). IEEE. https://doi.org/10.1109/VR.2014.6802080

Turk, M.A. (1991). *Interactive-Time Vision: Face Recognition as a Visual Behavior*: Massachusetts Institute of Technology.

Turner, J. H. (1988). *A theory of social interaction*. Stanford, Calif.: Stanford University Press.

Uddin, M. Z., Thang, N. D., & Kim, T. S. (2016). Human Activity Recognition Using Spatiotemporal 3-D Body Joint Features with Hidden Markov Models. *KSII Transactions on Internet and Information Systems, 10*(6). https://doi.org/10.3837/tiis.2016.06.017

Universität Augsburg University. (2017). EmoVoice: EmoVoice - Real-time emotion recognition from speech. Retrieved from https://www.informatik.uni-augsburg.de/en/chairs/hcm/projects/tools/emovoice/

Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using Presence Questionnaires in Reality. *Presence: Teleoperators and Virtual Environments*, *9*(5), 497–503. https://doi.org/10.1162/105474600566989

Usui, H., & Omura, Y. (2007). *Advanced methods for space simulations*. Tōkyō: TERRAPUB.

UTS:HELPS. (2009). Higher education language & presentation support.

Vaus, D. A. de. (2001). *Research design in social research*. London: SAGE.

Vinayagamoorthy, V., Steed, A., & Slater, M. (2005). Building characters: Lessons drawn from virtual environments. *Proceedings of the CogSci-2005 Workshop, 'Toward Social Mechanisms of Android Science'*, 119–126.

Vogel-Walcutt, J. J., Fiorella, L., Carper, T., & Schatz, S. (2012). The Definition, Assessment, and Mitigation of State Boredom Within Educational Settings: A Comprehensive Review. *Educational Psychology Review*, *24*(1), 89–111. https://doi.org/10.1007/s10648-011-9182-7

Waly, A. F., & Thabet, W. Y. (2002). A Virtual construction environment for reconstruction planning. (12), 139–154.

West, R. L., & Turner, L. H. (2017). *IPC3: Interpersonal Communication* ([3rd edition]). Boston, MA: 4LTR Press / Cengage Learning.

Wiederhold, B. K., & Bouchard, S. (2014). *Advances in virtual reality and anxiety disorders*. *Series in Anxiety and Related Disorders*. New York: Springer.

Wiederhold, B. K., Jang, D. P., Gevirtz, R. G., Kim, S. I., Kim, I. Y., & Wiederhold, M. D. (2002). The treatment of fear of flying: A controlled study of imaginal and virtual reality graded exposure therapy. *IEEE Transactions on Information Technology in Biomedicine*, *6*(3), 218–223. https://doi.org/10.1109/TITB.2002.802378

Wiederhold, B. K., & Wiederhold, M. D. (1998). A Review of Virtual Reality as a Psychotherapeutic Tool. *Cyberpsychology & Behavior : the Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, *1*(1), 45–52. https://doi.org/10.1089/cpb.1998.1.45

Wiederhold, B. K., & Wiederhold, M. D. (2005a). Specific Phobias and Social Phobia. In B. K. Wiederhold & M. D. Wiederhold (Eds.), *Virtual reality therapy for anxiety disorders: Advances in evaluation and treatment* (pp. 125–138). Washington: American Psychological Association. https://doi.org/10.1037/10858-011

Wiederhold, B. K., & Wiederhold, M. D. (Eds.). (2005b). *Virtual reality therapy for anxiety disorders: Advances in evaluation and treatment*. Washington: American Psychological Association.

Wilson, K., & Korn, J. H. (2007). Attention during Lectures: Beyond Ten Minutes. *Teaching of Psychology*, *34*(2), 85–89. https://doi.org/10.1177/009862830703400202

Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686

Wysoski, S. G., Lamar, M. V., Kuroyanagi, S., & Iwata, A. (2002). A Rotation Invariant Approach On Static - Gesture Recognition Using Boundary Histograms And

Neural Networks. *IEEE Proceedings of the 9th International Conference on Neural Information Processing, Singapura*.

Xu, P. (2017). A Real-time Hand Gesture Recognition and Human-Computer Interaction System. Retrieved from https://arxiv.org/pdf/1704.07296.pdf

Yamamoto, S., Yoshitomi, Y., Tabuse, M., Kushida, K., & Asada, T. (2017). Recognition of a Baby's Emotional Cry towards Robotics Baby Caregiver. *International Journal of Advanced Robotic Systems, 10*(2), 86. https://doi.org/10.5772/55406

Yin, S., Yang, J., Qu, Y., Liu, W., Guo, Y., Liu, H., & Wei, D. (2018). Research on Gesture Recognition Technology of Data Glove Based on Joint Algorithm. In *Proceedings of the 2018 International Conference on Mechanical, Electronic, Control and Automation Engineering (MECAE 2018).* Paris, France: Atlantis Press. https://doi.org/10.2991/mecae-18.2018.8

Young, R. O. (2017). *Persuasive communication: How audiences decide* (Second Edition). New York NY, London: Routledge Taylor & Francis Group.

Youngblut, C. (2003). *Experience of Presence in Virtual Environments.*

Yu, F., Chang, E., Xu, Y.-Q., & Shum, H.-Y. (2001). Emotion Detection from Speech to Enrich Multimedia Content. *IEEE Pacific Rim Conference on Multimedia 2001*, 24–26.

Zanbaka, C. A., Ulinski, A. C., Goolkasian, P., & Hodges, L. F. (2007). Social responses to virtual humans. In M. B. Rosson & D. Gilmore (Eds.), *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07* (p. 1561). New York, New York, USA: ACM Press. https://doi.org/10.1145/1240624.1240861

Zeltzer, D. (1992). Autonomy, Interaction, and Presence. *Presence: Teleoperators and Virtual Environments*, *1*(1), 127–132. https://doi.org/10.1162/pres.1992.1.1.127

Appendix