Ana-Despina Tudor

Development of a virtual audience concept for public speaking training in immersive virtual environments

kommunikationswissenschaft
interdisziplinär [kw.interdisziplinär]

Herausgegeben von Prof. Dr. Nicola Döring
und Prof. Dr. Jens Wolling

Institut für Medien und Kommunikationswissenschaft
an der Technischen Universität Ilmenau

Band 5

Development of a
virtual audience concept for
public speaking training in
immersive virtual environments

Ana-Despina Tudor

Universitätsverlag Ilmenau
2016

Impressum

V

Acknowledgments

Contents

List of Figures

List of Tables

List of acronyms

χ^2	Chi square
ϕ	phi
3D	Three-dimensional
AI	Artificial intelligence
ADD	Attentional deficit disorder
BAP	Body Action and Posture
CAVE	Cave Automatic Virtual Environment
CBT	Cognitive behavioral treatment
CID	Communication in the disciplines
CXC	Communication across the curriculum
CST	Communication skills training
df	degrees of freedom
GUI	Graphic user interface
ECA	Embodied conversational agent
EVE	Educational virtual environment
HCI	Human-computer interaction
HMD	Head-mounted display
M	Mean
MUD	Multi-user dungeon
n	number of cases
NCA	The National Communication Association
OCC	Ortony Clore and Collins model of emotion
PTSD	Post-traumatic stress disorder
Q&A	Question and answer
QCA	Qualitative content analysis
PI	Place illusion
Psi	Plausibility illusion
REA	Real estate agent
SPECIES	Special purpose embodied conversational intelligence with environmental sensors
SD	Standard deviation
SKANS	Seated Kinesic Activity Notation System
TE	Training expert
UCD	User-centered design

VE	Virtual environment
VH	Virtual human
VLE	Virtual learning environment
VP	Virtual patient
VR	Virtual reality
VRET	Virtual reality exposure therapy
VRTechE	Virtual reality technology expert

Abstract

The present dissertation proposes a novel approach to public speaking training and introduces a concept for a virtual audience in immersive virtual environments. The key component of such an audience are the virtual humans (VHs) whose role is to look and behave as closely as possible to a real public. Virtual audiences can then be used as training tools for people who wish to improve their public speaking and presentation skills.

Two empirical studies were conducted to help identify relevant audience features and behaviors that occur in real life and that can be modeled for a virtual public.

The first one is an exploratory study with experts in communication training and in fields related to virtual reality (VR). Seven experts were interviewed on the role and importance of audience customization, interaction, and group dynamic during training, as well as on technical possibilities to design virtual audiences with such features. The results show that audiences for communication training require extensive customization and interactivity options. To complement the findings, a speech practice session in a communication training seminar was observed and helped reveal the role of speech practice in the economy of the whole training seminar as well as the preferred feedback methods.

In the second study, a video observation of a student audience during a lecture at Technische Universität Ilmenau was conducted. The nonverbal behaviors of 14 students were coded and divided into attentive and inattentive manifestations. The identified behaviors are further described and analyzed in terms of their frequency, complexity, and duration.

The findings of both these studies helped create a design concept for a virtual audience with various characteristics (e.g., demographic features and virtual spaces they could inhabit) and a list with attentive and inattentive nonverbal behaviors the virtual humans could display. A five-minute scenario with virtual listeners is suggested as well.

1

Introduction

> *It usually takes me more than three weeks to prepare a good impromptu speech.*

Mark Twain

Giving a presentation at a conference, speaking in front of fellow students at university, presenting in front of managers–these are some of the usual speech situations people have to face at work or during their study. In spite of knowing that sooner or later people have to give a presentation or a speech and in spite of preparing it, the fear of being judged by others, of forgetting what was there to be said, and ultimately of failure can still lurk in the back of the mind. It's no surprise that people rank fear of public speaking very high on their list of fears. In a recent study, students were asked to rank their top fears, and public speaking reached the second position right after death (Dwyer & Davidson, 2012). In this context, it is justifiable to look at ways to help people overcome their anxieties, help them regain their self confidence, and improve speech performance.

Speech anxiety may be so intense that it interferes with one's ability and willingness to perform, and people have to undergo specialized treatment and learn how to lower their anxiety levels (Allen, Hunter, & Donohue, 1989). However, there are also people who do not require any treatment and who simply wish to learn and improve their presentation skills. For both phobics and non-phobics, skill improvement is a successful method to lower anxiety and improve performance (Fremouw & Zitter, 1978), and these could be the first groups to benefit from a training tool such as a virtual audience.

Compared to public speaking training, speech anxiety has already been extensively addressed in virtual reality (P. L. Anderson, Zimand, Hodges, & Rothbaum, 2005; Harris, Kemmerling, & North, 2002; Kothgassner et al., 2012; Pertaub, Slater, & Barker, 2002; Safir, Wallach, & Bar-Zvi, 2012; Wallach, Safir, & Bar-Zvi, 2009, 2011). Therefore, the research conducted here has a different focus on a less studied activity–public speaking and on the relevant topics of interest of a less researched stakeholder group–communication trainers.

The current research starts from previous work on anxiety treatment conducted in immersive virtual environments (IVEs) such as head

mounted displays (HMDs), CAVE Automatic Virtual Environments (CAVEs) or CAVE-like environments. Immersive virtual environments have supported numerous applications that addressed phobias, and they proved to be a viable alternative to *in vivo* exposure. In the case of fear of flying for instance, patients would consider exposure to an actual flight too extreme to actually fly and face their fears (Wiederhold et al., 2002). The thought of simulating a virtual audience to treat public speaking seems also more plausible than having real people act as one for the duration of the treatment. The attention toward virtual reality applications is further justified by the high realism offered by existing VR technologies, which allow user to experience immersion and presence within the simulated scenario (Slater, Linakis, Usoh, & Kooper, 1996). Previous applications in anxiety treatment have explored the possibilities and proven the success of virtual reality in addressing acrophobia (Emmelkamp et al., 2002), agoraphobia (Botella et al., 2007; Meyerbröker, Morina, Kerkhof, & Emmelkamp, 2011), arachnophobia (Garcia-Palacios, Hoffman, Carlin, Furness III, & Botella, 2002), and aviophobia (Rothbaum et al., 2006). Apart from addressing speech anxiety disorder, as in the case of glossophobia highlighted above, available technologies could transform a virtual audience into a viable training tool for presentation and public speaking skills. Communication skills training represents a field with high applicability potential and will be therefore further addressed.

Combining communication skills training with VR technology requires adaptation on both sides. Various studies on virtual reality exposure treatment (VRET) discussed in the coming chapters showed that VR technologies offer the opportunity of adaptable speech scenarios, beyond the usual treatment setting in real life. Within the scenarios, the characteristics of virtual humans that have been used for phobia treatment can be revised to satisfy the needs of non-phobic trainees, and can be customized further for those who wish to improve particular skills in public speaking, such as preparation for public speaking competitions. Other adjustments refer to the role of the trainers and their particular needs and expectations from a virtual reality training simulation adapted to their training groups. These would be examples on how technology adjusts to serve the training purpose. Furthermore, using VR technology within training sessions requires its appropriation. This is revealed in the way trainers envision using VR within their training sessions and organizing their training with and around the

technology. All these points will be approached in the first empirical study of the thesis.

Communication skills represent a key factor in the preparation for competent, effective, and responsible citizens, and many universities in the world have programs to teach students how to communicate effectively (Dunbar, Brooks, & Kubicka-Miller, 2006). Furthermore, communication skills trainings have been implemented lately across various disciplines and they prepare students and professionals to communicate effectively in their own field of practice and beyond, to broader audiences, such as potential clients (as in engineering and design fields) (Dannels & Housley Gaffney, 2009) or patients (as in the medical field) (Parle, Maguire, & Heaven, 1997). To address this variety of communication scopes, the virtual audience proposed here:

- is meant to be customizable and to contain virtual humans that can depict various demographic characteristics and outfits;

- is meant to be displayed in different virtual locations, depending on the needs of the users;

- is active and can depict various behaviors, such as being attentive or inattentive;

- can accommodate sensors to foster interaction with the user.

The group of virtual listeners embodying these characteristics would then allow trainers to chose among various behaviors and 3D immersive training environments that fit best the needs of public speaking trainees.

CHAPTER 1.1

Relevance of research

Virtual audiences (VAs) are common in anxiety treatment studies, as those mentioned above, and the dissertation builds upon the existing knowledge in order to expand VAs' applicability outside a therapist's office. This aspect is best revealed by moving the attention towards a different stakeholder group than before–public speaking and presentation trainers and trainees. Trainees are a broad group of people with various levels of anxiety. Some are phobics and require treatment, some are not and require only skill training for specific public speaking or presentation events. It can be argued that the novelty of the research conducted for the dissertation lies in the integration, within the design, of new stakeholder groups and their views. Moreover, new audience characteristics are analyzed and proposed in the expert interviews. Further on, the findings of the audience observation underline the multimodal displays of behaviors which have been extensively researched with embodied conversational agents in one-to-one interaction with users, but not with virtual groups. As Kang et al. (2013) pointed out, there are few autonomous audiences as well as few ones used outside anxiety treatment.

Lastly, the virtual audience concept proposed here highlights the elements that support skill transfer from the virtual environment into the real world and rests on a theoretical foundation of virtual learning environments design. In the following chapters several study domains are combined to address theoretical and empirical gaps in the field and offer an integrative virtual audience concept.

CHAPTER 1.2

Chapter overview

Following the general introduction, the dissertation is structured into five parts:

- an introductory part that addresses the audience research model which frames the inquiry;

- a part that reports the first empirical study on audience characteristics;

- a part that reports the second empirical study on audience behaviors;

- a part that presents the virtual audience design concept proposed for communication skills training and a virtual audience scenario;

- a final part covering a general discussion.

The introductory part addresses communication skills and existing training methods. Special attention is given to the concept of communication in the discipline, as one of the main trends in the communication skills research, which pleads for a customization of training in line with the values, norms, and epistemologies of that discipline (Dannels, 2002). The chapter also points out the difference between anxiety treatment and skills training. Next, several VR-related concepts are presented, followed by speech anxiety treatment in virtual reality. Following, there is a brief review of virtual humans' characteristics and the behaviors they can display, especially in the case of virtual groups. The last chapter draws on the literature researched in the previous chapters and introduces the virtual audience research concept.

The next two parts of the dissertation contain two empirical studies. The first study presents in-depth interviews with public speaking

trainers and virtual reality application technology experts. These experts discuss the importance of feedback and interaction between speaker and audience, as well as ways to customize and animate virtual audiences to act like unitary groups. The findings of the interviews are supported by a participant observation of a speech practice session within a three-day presentation seminar held at Technische Universität Ilmenau. The results of the empirical studies helped identify the role of audience diversity, interactivity between speakers and listeners, as well as the role of dynamic audiences in public speaking. Moreover, they uncovered additional data on audience demographic characteristics, their behaviors, training practices, and trainers' expectations from a virtual audience. Further on, technology experts proposed technical solutions to what characteristics can be implemented for the virtual audience.

The second empirical study presents the observation of a real audience focused on the nonverbal behaviors manifested by listeners. The sample contains 14 students that were observed for a total of ten minutes each, at the beginning, middle, and end of a lecture. Behaviors were coded and analyzed in a multimodal fashion. The analysis identified what nonverbal behaviors occurred simultaneously and which of them engaged several modalities. It was possible to identify what people do with their hands, torso, heads, and gaze while listening to the presenter. To make behavior interpretation easy, two main interpretation frames were given, and behaviors could be assigned to either one of them: an attentive behavior or an inattentive one. Furthermore, understanding the multimodal nature of nonverbal behavior display was central to the analysis. Multimodality provides several rules for animating virtual humans in the future, by revealing through the data analysis what body parts are simultaneously active during a certain time frame.

The findings of the empirical studies serve as background for the virtual audience design concept and the virtual audience scenario presented in the fifth part. Following the theories and the design prerequisites of virtual environments design in education, the design concept and scenario offer a thorough list of audience characteristics, behaviors, and interface features that can be implemented for public speaking in VR. The dissertation closes with a general discussion which reviews all findings and discusses general limitations and opportunities for future work in the virtual training domain.

2

Audience research model for communication training in virtual reality

CHAPTER 2.1

Communication skills

2.1.1 Introduction

The design of an audience concept for public speaking and presentation training in virtual reality is a multidisciplinary endeavor at the intersection of communication skills training, speech anxiety treatment in virtual reality, and virtual reality technologies.

On the one hand, there is the educational perspective on teaching communication skills (Beebe, 2007) that reviews methods of training and evaluation of skill acquisition. Communication skills address a general population and are usually taught in university classes or in separate seminars and workshops during study time or at work (Beebe, 2007; Furbay, 1965; Turk, 2004). On the other hand, there is the psychology angle that focuses on public speaking phobia and on treatment methods. Compared to communication training which is only now starting to benefit from VR technologies (e.g., doctor-patient communication (Stevens et al., 2006)), fear of public speaking has been extensively studied in virtual environments (a full review is given in the next chapter), and several treatment solutions can already be found on the market (e.g., Virtually Better).

The present chapter will introduce concepts such as communication skills, public speaking, and presentations which form the starting point for the creation of the virtual audience design concept proposed here. The literature review is meant to build the context around virtual reality as tool for public speaking and presentation skills training.

2.1.2 Oral communication styles

A speech genre is part of the broader discourse realm and is defined as "each sphere in which language is used [...] and where it develops its relatively stable types of [...] utterances" (Bakthin, 1986, p. 60). Speech genres include various communication types such as public speaking, interpersonal or informal speaking, but, for instance, also discussing literature or presenting results of an engineering project (Darling & Dannels, 2003).

There is a rich terminology used in literature that describes giving a speech in public (Dance, 1987). This type of activity falls intuitively under the generic term of communication skills. However, scholars have debated whether there are notable differences between giving a presentation and giving a public speech depending on context of speech and on genre (P. Rogers, 1988). Dance (1987) argues that the term "presentation" has subtly replaced the term "public speaking" in the business field. Moreover, he states that there is only a difference of context between the two practices and no difference of genre between the two terms–one takes place in front of a few people like in a business pitch, the other in front of more people gathered in a public space.

Furthermore, P. Rogers (1988) considers that presentations and public speeches differ also at genre level, since they imply a different type of interaction with the audience. Quoting several studies (Howell and Bormann, 1971; Thrash, Shelby, and Tarver, 1984), she argues that presentations make use of visual aids, are delivered extemporaneously, are interactional, are detailed, and can have more than one presenter (e.g., sales pitches, progress reports, briefing sessions, explanations of policies, plans, and proposals). Therefore, they have greater dialogic value than a public speech, as listeners can ask questions at any time. On the other hand, public speaking as genre includes formal introductions, press statements, or presidential addresses (P. Rogers, 1988). However, for simplicity reasons, both presentations and public speaking will be used here interchangeably.

Apart from this, one can separate between several public speaking styles, and these are differently ranked in importance by trainers and speakers. In one study, 267 public speaking faculty members and 179 alumni ordered delivery styles by relevance in the public speaking course and at work (Johnson & Szczupakiewicz, 1987). Public

speaking instructors considered extemporaneous delivery style (with preparation) as the most important, followed by impromptu delivery style (without preparation). Alumni raked them the other way around, and considered the impromptu delivery style the most important one both in the basic public speaking course and in the workplace. The same study analyzed the types of public speaking skills taught in the classroom and how these are actually used in the workplace. The most often taught skills in faculties were speech begin and ending techniques, discourse body organization, and methods for supporting materials collection. Alumni however, needed to use different skills at work, such as giving informative speeches, handling questions and answers, and listening. Among the least used skills by alumni were evaluating audiences and evaluating speeches (Johnson & Szczupakiewicz, 1987). This differentiation represents an opportunity for the training tool proposed here, as it could be adapted to meet the needs of both trainers and trainees.

2.1.3 Communication skills definition

Communication skills training (CST) has gained a lot of popularity in the past decades and has become an intensively researched field (Street & De Haes, 2013). Communication skills are part of the broader social skills domain and can be traced back to a technique called "microteaching", which was applied to teachers in order to improve their teaching skills in the 1960s (Hargie, 2006). According to the author, the main quality of the microteaching technique was that the addressed skills were broken down into smaller components (e.g., knowing how to ask questions). These could be taught separately across different stages, such as the preparation stage, the training stage, and the evaluation of the program stage (Hargie, 2006). In spite of its apparent advantages, this approach has also received criticisms for its atomizing view on communication, which, according to Gestalt views, cannot be broken down into simple, hard to define components such as smaller and separate skills (Hargie, 2006; Salmon & Young, 2011). Nevertheless, there are advantages to microteaching reflected in the current organization of communication skills training seminars that focus on specific skills (e.g., giving and receiving feedback), as seen in the expert interview and the observed speech training session described in later chapters.

Overall communication competences include multiple skills such as speaking, listening, cultural sensitivity in communication, understanding values and ethics, or critical thinking (Morreale, Rubin, & Jones, 1998). As pointed out by Quianthy (as cited in Morreale et al., 1998) the National Communication Association (NCA) in the US identifies a competent speaker as a person who is "able to compose a message and provide ideas and information suitable to the topic, purpose, and audience" (p. 7). The way these messages are conveyed and delivered depends on factors such as topic understanding, information organization, and strategic usage of verbal and nonverbal communication channels (Morreale et al., 1998). The targeted verbal communication skills cover the content of the delivery, such as word choice and argumentation, but also delivery style, such as voice pitch, or voice intonation (Morreale et al., 1998). Matching the verbal intent with the nonverbal manifestation is achieved through a set of expected nonverbal skills (Morreale et al., 1998):

- usage of appropriate paralanguage such as voice tone or emphasis;
- usage of appropriate kinesics such as facial expressions, posture, and gestures;
- usage of appropriate proxemics, by managing interpersonal distance properly;
- usage of appropriate apparel and ornamentation.

The NCA has devised speaking and listening competencies that should be taught to undergraduate and graduate students (Morreale et al., 1998), as well as instruments for assessing skill acquisition (Morreale, Moore, Surges-Tatum, & Webster, 2007). The Competent Speaker Evaluation Form assesses eight public speaking skills from unsatisfactory to excellent:

- topic choice and adaptation to audience and occasion;
- appropriate communication of the thesis or speech purpose to audience and occasion;
- usage of supporting materials (e.g., presentation slides);
- appropriate information organization for the speech purpose, audience, and occasion;

- appropriate language usage;

- appropriate usage of paralanguage elements such as voice pitch and intensity to keep audiences interested in the speech;

- proper usage of pronunciation, grammar, and articulation;

- usage of nonverbal communication that complements the verbal message.

The list of competencies and skills evaluation elements offers a good starting points for planning audiences and scenarios in virtual reality to address communication competence. This is however a normative approach that would have to be considered in the light of current technical possibilities offered by state of the art VR technologies. A fair question follows: For what kind of competencies can a virtual audience train, by giving feedback to? To answer this question, all skills would have to be broken down into precise measurable components and matched with existing technologies. For instance, voice pitch and intensity could be picked by sensors, which in turn allow a virtual audience to respond to changes of these variables by acting out certain behaviors. Skills such as appropriateness of language usage would require a match with a language processing systems. The empirical studies presented in the upcoming chapters explore the training needs and the possibilities to devise interactive virtual audiences during training sessions that respond to nonverbal communication alone.

2.1.4 Communication skills training

One can speak broadly of two types of communication training, which have been reviewed in the literature: the basic communication course and the advanced speaking course. Morreale, Hanna, Berko, and Gibson (1999) define the first type of course as the one required or recommended for the majority of the undergraduate students. Whereas the basic course serves as an introduction to speaking in public, the advanced speaking course addresses speakers who wish to improve their experience beyond basic knowledge and reach expert levels (Levasseur, Dean, & Pfaff, 2004). The virtual reality application for public speaking and presentation skills proposed here is meant to address both the basic and the advanced participants.

According to Beebe (2007), communication skill training contains five general steps:

1. being told how to perform a skill;

2. being shown how to perform it;

3. being invited to perform;

4. being encouraged to perform and receiving feedback, and

5. having the performance corrected, in order to improve.

A review of training practices in the business sector conducted by Putnam (1979) revealed that trainers prefer to assess trainees' needs before devising the training content. Moreover, even though lecture-discussions ranked on the first place among teaching methods, almost half the trainers incorporated role-playing sessions, media presentations, as well as case studies (Putnam, 1979). Whereas the literature on communication skills gives little account on communication skills training procedures, it focuses usually on the dos and don'ts on skill improvement from the speaker's perspective (Beebe & Beebe, 2010, 2012; Lucas, 2012). This is also one of the reasons why trainers were encouraged to discuss their training procedure in the first empirical study of the dissertation.

2.1.5 Communication within and across disciplines

In the English speaking area, in the past 40 years, the study of communication programs has been marked by increased attention towards expansion and adaptation to more academic disciplines, outside communication studies. This movement is called "communication across the curriculum" (CXC) and it refers to "the implementation of communication instruction in disciplines other than communication studies–typically in the form of a university program" (Dannels & Housley Gaffney, 2009, p. 125). This approach to communication skills training has been particularly researched in the US in fields such as design, engineering, and medicine (Dannels & Housley Gaffney, 2009).

It became clear in the academia that communication is a skill that any successful and competitive employees should possess, especially

in the fields where they interact with clients and have to translate their knowledge into a lay person's language (Dannels, 2002). This led further to a study framework called communication in the discipline (CID), which suggests that "communication is a situated, contextual activity" (Dannels & Housley Gaffney, 2009, p. 137) and should also be part of the university curricula. According to CID, communication skills should be adapted to each study discipline and integrate the norms, values, and epistemologies of that particular discipline (Dannels & Housley Gaffney, 2009). The CID framework is grounded in the situated learning theory (Dannels & Housley Gaffney, 2009) which makes communication a context-driven activity.

In this direction, health care students, for instance, learn how to deal with patients (Bartlett et al., 1984; Stevens et al., 2006) and how to communicate vital information about medication (Chant, Tim, Randle, Russell, & Webb, 2002). Health care professionals have sometimes difficulties with discussing sensitive health problems, break bad news, or challenge denial with their patients and these are necessary skills especially in the palliative care of cancer patients (Faulknera, Webbb, & Maguire, 1991; Parle et al., 1997). In mental health training, it is common that students train clinical assessment or interview skills (Kenny, Parsons, Gratch, Leuski, & Rizzo, 2007). During the training, students have to practice with actors who portray various physical or psychological problems, which have to be assessed and tackled accordingly by the trainee (Kenny et al., 2007). Actors represent a limited resource for dialogue and scenarios, and several studies have started researching doctor-virtual patient (VP) interaction, where VPs engage trainees in natural conversations (Bickmore & Giorgino, 2006).

Engineering students face different challenges, such as having to present their product in simple and understandable ways to potential clients, to lay persons, or to investors (Dannels, 2002). Their communication skills revolve around a general process of translation and adaptation of their expert knowledge into understandable data, through visual aids, using clear numbers, or focusing on results (Dannels, 2002). Apart from this, acquiring lecture (M. Hoffman & Mittelman, 2004) and conference presentation skills that cut across any study or research field (Vickers, 1997) have also come under the spotlight lately. This is reflected for instance in higher education communication skills training preparing students for seminar and conference presentations (Hughes & Large, 1993).

A virtual reality audience points out several opportunities for this kind of communication training. The CID paradigm speaks of adapting speech to a target audience and its characteristics. There are countless possible target audiences someone could train for and, as seen above, they can range from colleagues in the same field of work and potential clients to patients. Another advantage of virtual characters and virtual audiences in communication training and practice is their availability: one can use them for as many trainees as possible, for as long as necessary, and on various platforms (e.g., Justina, the virtual patient on a desktop computer (Kenny et al., 2007); DIANA, the virtual patient and VIC, the virtual instructor on a projection wall (Johnsen, Dickerson, Raij, Lok, & Jackson, 2005)).

2.1.6 Communication skills for state and trait apprehension

Public speaking has been presented in the previous sections as a skill that can be acquired and not as phobia that requires treatment. Speech anxiety is briefly introduced and its two types are explained and differentiated below.

There are two ways to look at fear of public speaking: from the perspective of state apprehension and of trait apprehension (McCroskey, 1977). Trait apprehension is the anxiety one feels independently of the speech or communication context. State apprehension is bound to a certain context, such as speaking at a conference or speaking in front of certain audiences (McCroskey, 1977). The university courses on communication skills tend to overlook causes of anxiety (trait anxiety) (McCroskey, 1977) and to focus on state apprehension, by training presentation skills for seminars, conferences, and for future work situations (Beebe, 2007; Turk, 2004). However, trait anxiety can also occur among those people that have to undergo communication skills training as part of their university curricula, and researchers argue that what is typical training for state communication apprehension can have different effects on people who suffer from trait communication apprehension (McCroskey, 1977). Such training could actually reinforce the apprehension to communicate, therefore participants could avoid speaking in public during communication classes even more and feel highly uncomfortable if pressed to do so (McCroskey, 1977).

Speech apprehension may be hard to spot and screening before public speaking courses could be useful for identifying potentially communication apprehensive people. This differentiation between state and trait apprehension is relevant for communication skills training beyond its procedural implication of screening participants. It supports one of the main ideas of the audience concept, that of customizing the training tool to the needs of various users and of offering solutions with various degrees of interaction and exposure to speech situations.

Communication skills training programs for state apprehension are organized around many foci that vary from basic skills such as delivery style and argumentation to advanced skills that target greater levels of behavioral adaptability (Morreale & Backlund, 2002). Independently of how complex the training is, there is always a theoretical part which walks the trainee through all necessary skills they need to acquire followed by a practical part where they practice their newly learned speech skills (Verderber et al., 2008). The next section discusses such communication skills training techniques applied to non-phobic trainees.

2.1.7 Summary

Attaining proper communication skills for work and study purposes requires a good understanding of many factors, such as purpose of communication, speech context, audience profiles, information structure, etc. The chapter introduced some of the current topics in the field and showed that sustained effort is taking place to prepare people for an ever growing diverse audience in their study and work environments. Skilled communicators are expected to posses a whole array of skills that are taught in communication training classes, and interactive virtual characters can teach physicians how to address patients and their families. In this context, a virtual audience for speech and presentation skills training appears as a sensible step to take.

CHAPTER 2.2

Virtual environments and applications in communication training and treatment

2.2.1 Introduction

The previous chapter mentioned a few studies that use virtual reality systems in training communication skills to medical students (e.g., in Stevens et al. (2006)). This chapter introduces and defines the main concepts and characteristics of virtual reality and virtual environments and points out the opportunities and risks they offer in relation to the creation of a virtual audience for communication training. The focus here lies mainly on technology and the chapter serves as bridge between communication skills training (the driving motive of the audience concept), VR as platform for training, and virtual audiences used in the related field of speech phobia treatment. The latter ones are reviewed in Section 2.2.5 of the present chapter.

In 1987 Jaron Lanier coined the term "virtual reality" for something that he, as he explained in an interview, couldn't describe better with other terms, such as synthetic, telereality, cyberspace, or shared dream (Lanier, 1989). He went on saying that "virtual reality is not a computer. We are speaking about a technology that uses computerized clothing to synthesize shared reality. [...] It only has to do with what your sense organs perceive" (p. 110). With the advance of technology, definitions of virtual reality tended to be limited to newly introduced and used devices. Fuchs, Moreau, and Guitton (2011) criticized this simplistic view and grasped its complexities through a multi-leveled definition: by its purpose, its functions and its technical affordance: "Virtual reality is a scientific and technical domain that uses computer science and behavioral *interfaces* to simulate in a virtual world the behavior of 3D entities, which interact in *real time* with each other and with one or more users in pseudo-natural *immersion* via sensorimotor channels" (italics by author) (Fuchs et al., 2011, p. 8). Interfaces serve

for communication from the human user to the computer and vice versa, by informing the user about the changes in the virtual scenery through visual, auditive or haptic stimuli and by informing the system about the movements of the user via sensors. Real time interactivity ensures that the system is responsive within acceptable latency levels without breaking "the magic circle" of the simulation (Fuchs et al., 2011). The authors also qualify immersion as a pseudo-natural occurrence, arguing that virtual reality creates perception biases and humans naturally notice differences between reality and virtuality. Immersion is one of the most pervasive concepts in virtual reality research and will be addressed separately in the following sections.

Virtual reality has often been used interchangeably with virtual environments (VEs) (Burdea & Coiffet, 2003), and the definition given by Blascovich et al. (2002) to VEs bears similarities to the previous one above: VE represents "synthetic sensory information that leads to perceptions of environments and their contents as if they were not synthetic" (p. 105).

Virtual environments can address all sensorial perceptors of the human body, from visual, auditory, and haptic to olfactory, and gustatory, and allow users to move (Blascovich et al., 2002), as well as travel through and interact with the depicted virtual objects (Bowman, 1998; Bowman, Gabbard, & Hix, 2002) or characters (Gratch et al., 2002). Interaction with a virtual audience could theoretically involve all these senses, as long as this serves the training purpose. However, for the audience concept proposed here, only the visual and the auditory inputs will be considered, as these are the most relevant in the communication process of public speaking.

2.2.2   Immersive virtual environments

There are mainly two types of devices that help users detach from the surrounding real environment and immerse into a 3D synthetic world while still perceiving it as realistic: head-mounted displays, such as for instance Oculus Rift and room-sized environments such as CAVE-like and CAVE Automatic Virtual Environments (Blascovich et al., 2002). Through high technical fidelity images (e.g., high resolution, low latency), users become immersed and eventually ought to experience

presence or the felling of actually being within the simulated virtual world (Slater, 2009).

HMDs are a helmet-like device with a display that is capable of showing 3D images (Shibata, 2002). Some of the early models had to face issues such as graphic quality, weight, and mobility, due to cabling (Shibata, 2002). Currently there are models that use smart phones encapsulated in cardboard frames (e.g., Google Cardboard (Google, 2016)). Compared to the HMD where sharing the same virtual reality simulation is possible only by using separate devices, a CAVE Automatic Virtual Environment allows multiple users to share the experience in the same physical space.

The CAVE is "a multiperson, roomsized, high-resolution, 3D video, and audio environment" (Waly & Thabet, 2002, p. 151). There are several configurations, with three (lateral) or six (lateral, ceiling, and floor) mobile walls. The lateral walls can shift from a common wall into a cube-shaped environment (walls are at 90 degrees relative to each other) (see Figure 2.1). 3D images can be projected on all walls and can be viewed with shutter glasses or polarized glasses (Waly & Thabet, 2002). When a user wears position tracking sensors, the perspective of the images changes along with the position of the user, allowing the user to always see the correct perspective in real time from his/her current position. Several people can go simultaneously into the CAVE or the CAVE-like environment, but the stereo images will adapt only to the position of the person wearing tracking sensors (Beier, 2008).

The virtual audience concept is proposed for a CAVE-like virtual environment with three frontal projection walls (2,8 x 2,1 m each), a total length of 8,4 m and a total height of 3,24 m, as seen in Figure 2.1. The gap between the walls is less than 1mm (FASP, 2014). The device is located on the campus of Technische Universität Ilmenau, Germany.

One of the first applications used the CAVE in the visualization of the universe, its expansion, and colliding galaxies (Cruz-Neira, Sandin, & Defanti, 1993). Other early applications visualized fractal exploration and strange attractor formation (Cruz-Neira et al., 1993). CAVEs have been used also in architecture for walk-throughs, where users and architects can explore the virtual interiors of houses or of historic sites (Cruz-Neira et al., 1993).

Figure 2.1. Illustration of the CAVE-like immersive virtual environment in the FASP Laboratory at Technische Universität Ilmenau (FASP, 2014)

Another application is virtual prototyping. In a study by Seron, Gutierrez, Magallon, Sobreviela, and Gutierrez (2004) 600 users could see a train prototype and make suggestions on how to improve seat design and colors. Dunston, Arns, and McGlothlin (2007) used the CAVE in the design of a virtual hospital patient room. Users could see if there were any problems with how furniture was positioned or if there was enough space to move around. Nurses were particularly satisfied with the 3D life-size simulation and the feeling of immersion they experienced in the CAVE as well as with the quick image rendering they experienced while moving in the virtual environment. Another room design application required both patients and medical personnel to evaluate hospital wards, and findings showed that the CAVE environment could help evaluate factors such as furniture arrangement and room lighting. Based on these examples, some of the advantaged of the CAVE emerged: its high level of immersion compared to desktop simulations (Dunston et al., 2007), low cost simulation compared to real prototyping (Seron et al., 2004), and possibility of large data visualizations (Cruz-Neira et al., 1993).

Immersive virtual environments (IVEs) prove useful also from a psychological perspective, for their role as research tools. They are particularly appealing because they enable keeping the experimental control as well as allow a relatively high mundane realism (Blascovich et al., 2002). Researchers are able to control the simulation and the simulation technology can reach a high degree of mundane realism (Figure 2.2.C). By comparing traditional methods to multimedia methods and immersive virtual environments, the authors argue that IVEs have great potential at keeping the ratio between high realism and high experimental setting (Blascovich et al., 2002).
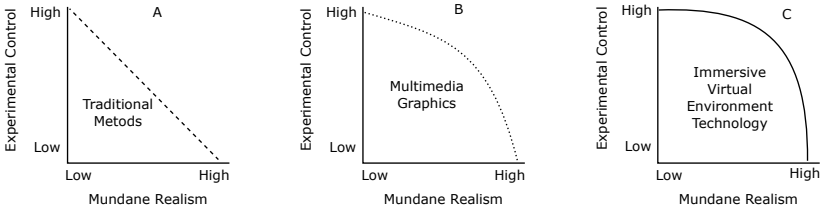
Figure 2.2. Types of realism (Blascovich et al., 2002, p. 104)

Due to these qualities, and to the advances in animations and realistic displays of virtual humans and objects, IVEs can be used successfully in various psychological investigations, such as visual perception studies, spatial cognition, proxemics, or social-facilitation inhibition, as well as tools for phobia treatment, such as acrophobia or arachnophobia (Blascovich et al., 2002). The next section will introduce two of the main concepts connected to IVEs–immersion and presence.

2.2.3   Concepts and definitions

There are two major intensively researched concepts that characterize immersive virtual environments: immersion and presence.

The quality of being immersive translates into the fact that users can be physically surrounded by the 3D projected images on the walls around them, as in the case of CAVEs and CAVE-like environments (Sherman & Craig, 2003) or wear a device that shuts out any sensorial input from the outer world, as in the case of HMDs (Shibata, 2002). Immersion is a complex concept that bears various definitions. Slater and Wilbur (1997) define immersion as the extent to which displayed images recreate a vivid illusion of reality, from a technical perspective. They suggest four features of immersion such as inclusiveness (how much physical reality is shut out), extensiveness (number of modalities addressed by the simulation, such as visual and auditory), surrounding (the shape of the walls, e.g., panoramic or cube-like, see Figure 2.1, and vividness or quality of the display (e.g., image resolution). The authors propose that each of these elements is scalable, and at its highest level of immersion a virtual environment shuts out the real world completely. Following this argument, one can speak of levels of immersion, rather than of immersive or nonimmersive environments (Wang & Bowman,

2007). Since the degrees of immersion depend on the property of a system, they ought to be measured objectively and independently of a human's experience (Sanchez-Vives & Slater, 2005). Furthermore, in an immersive system, the user is able to perform actions as if they were in physical reality: the simulation responds to the movements and actions of the user just as if these were in a physical environment and it allows users to perform actions just like in a real one (e.g., move around, grab objects, and have visual and haptic responses from the system) (Slater, 2009).

Whereas the former definition of immersion combines display quality with virtual object interaction in virtual reality, Bowman and McMahan (2007) use the term fidelity. They define immersion as "the objective level of sensory fidelity a VR system provides" (p. 38). (McMahan, Bowman, Zielinski, & Brady, 2012) continue with the term "fidelity", but distinguish between display fidelity and interaction fidelity. Experiments showed that both high display (CAVE) and high interaction fidelity (with a six degrees of freedom wand) had a positive effect on experienced presence, engagement, and usability when playing a game that required aiming, shooting, and movement within the simulated game world (McMahan et al., 2012).

Whereas researchers agree that immersion is a property of the system and of its technical affordances, presence is a more nuanced concept and there are various views on what the term defines (G. Mantovani & Riva, 1999).

Riva (2000) and G. Mantovani and Riva (1999) suggest that a different ontology leads to various meanings of presence, depending on the understanding of the reality concept. One can use three broad explanatory frames for presence in mediated environments–the ingenuous realism, the ecological perspective, and the cultural view (Ijsselsteijn & Riva, 2003; G. Mantovani & Riva, 1999). According to ingenuous realism, human knowledge perceives a preexisting reality, outside of the subject, and the existence of objects is real as long as they are located in a physical environment G. Mantovani and Riva (1999, p. 6). This poses problems for virtual reality, because objects depicted there don't have the physical existence in the acceptance of this view, yet they create perceptions in the human mind, and have a sort of "imagined" existence (G. Mantovani & Riva, 1999). They criticize this view and further suggest that this dichotomy is unfounded, and that

reality is constructed and culturally mediated and not dependent of technical mediation or lack thereof (G. Mantovani & Riva, 1999; Rettie, 2004).

In the view of the ecological approach, there is no reality independent of the subject, as in the previous view, and this reality is constructed by subjects through the affordances it provides. Accordingly, virtual reality should not only reconstruct spaces and sounds with high fidelity to create a feeling of presence, but also allow actors to take actions in the simulated environment and to fulfill their tasks (Ijsselsteijn & Riva, 2003; G. Mantovani & Riva, 1999).

The cultural approach states that presence is socially constructed and that faithful reproduction of physical reality in a virtual environment is important, yet not sufficient compared to creating a context where actors act and function freely in an ecological way (Mantovani, 1996, as cited in Ijsselsteijn and Riva (2003); (G. Mantovani & Riva, 1999)). In this view, presence is a feeling that depends on the context and content of the experienced environment and can be felt in less immersive environments as well (Ijsselsteijn & Riva, 2003).

These views offer a rich background for conceptualizing and further understanding presence, however, a definition analysis based on these three categories would go beyond the scope of this chapter. There is a common understanding of presence, as feeling of "being there", within a simulated place, in the work of Biocca (1997), Steuer (1992), and Witmer and Singer (1998), and which can be traced back to the telepresence concept (K. M. Lee, 2004). Lombard and Ditton (1997) define presence as "the perceptual illusion of nonmediation" (para. 33) and they propose six conceptualizations of presence: realism, immersion, transportation, social richness, social actor within medium, and medium as social actor.

K. M. Lee (2004) analyzes various definitions and develops his own one for presence as "a psychological state in which virtual objects are experienced as actual objects in either sensory or nonsensory way" (p. 37). Social presence is experienced when social actors are perceived as actual actors and artificial intelligence is successfully simulated. Most importantly, the users do not notice the artificiality of objects, social actors, and of selves. It remains to be seen though what happens when users still notices this and what effect their realization has on the effectiveness of task performance in virtual environments.

Social presence is experienced by people who feel they share together a common space (e.g., by using online chat or multi-user dungeons (MUDs)) (Ijsselsteijn & Riva, 2003). At the intersection between physical and social presence is co-presence–the experience of being together and sharing the same space (Ijsselsteijn & Riva, 2003). The latter is experienced when people negotiate their interactions in the shared environment (Ijsselsteijn & Riva, 2003).

Slater et al. (1996) and Slater (2009) define presence as a psychological response when (1) users experience immersion, (2) the virtual environment becomes dominant, and when (3) they recall having visited a place rather than having seen just images. Slater (2009) uses the term "place illusion" (PI) to define the feeling when an immersive system allows users to perform the same physical actions as those they would perform in reality while exploring the environment through their senses. Place illusion exists as long as the user doesn't approach to touch a CAVE's walls or see the pixels (Slater, 2009). A second phenomenon proposed by the author and which could be simultaneously experienced with place illusion is plausibility illusion (Psi). Psi refers to the experienced belief that what is happening in the simulated world is happening for real (Slater et al., 1996; Slater, 2009). The author argues that plausibility illusion is supported by the credibility degree of the events depicted in the simulation or how likely they are to occur. Plausibility is similar to realism and Blascovich et al. (2002) affirm that virtual characters are realistic when they behave similarly to the way they would in the physical world. To achieve this level of realism, they should be able to encode, decode, and interpret both verbal and nonverbal behaviors of fellow virtual characters with whom they share the virtual space as well as those of user avatars (Blanchard & Markus, 2002).

Social presence in particular is central in the discussion of virtual reality applications with virtual characters and (Biocca, Harms, & Burgoon, 2003) call for a new theory and definition to the concept. They classify the existing definitions into three categories, each highlighting a different aspect of the concept as reflected in existing theory. Co-presence definitions highlight mutual awareness of the other, the feeling of sharing the same space, the feeling of proximity of the other, and the possibility to interact. Psychological involvement definitions focus on perceiving the presence of another intelligence, its salience in the communication act, as well as a sense of immediacy and intimacy with the otherness. The last classification is behavioral engagement and, in

a way is the closest to the interpersonal communication in real life, as it pleads for virtual environments that allow the richest manifestation of behaviors through multichannel communication (Biocca et al., 2003). Although with different foci, all definition categories contain elements that point out towards interactivity with the "otherness" in the shared environment. Nevertheless, the authors suggest that a new unifying and broader definition ought to be developed to accommodate various interaction kinds over diverse media, with both human and artificial intelligence, and various levels of social interactions.

It is worth noting that presence definitions vary depending on the field where they've emerged and are malleable to adapt to various research domains (K. M. Lee, 2004). In early applications of virtual reality where interaction with the environment was focused on accessing work spaces from distance, achieving telepresence or bringing the user to that location was highly relevant (Minsky, 1980, as cited in K. M. Lee (2004)). When collaboration comes into discussion, social presence takes the stage (Biocca et al., 2003). F. Mantovani and Castelnuovo (2003) discuss this variety and argue that there's a need for new models of presence to accommodate an ever changing field. They propose a broad model of presence to address a new class of virtual reality applications for training purposes. They state that, in order to achieve a first had experience of the relevant skills and knowledge, virtual environments ought to "seem real and engaging to participants" F. Mantovani and Castelnuovo (2003, p. 168). They reorganize presence factors across the field into factors applicable to virtual training environments. At perceptual levels, there is graphic fidelity (similar to immersion) and user's control over the simulation through objects manipulation and feedback from virtual characters. The authors speak in general terms about interactivity with the virtual content both in physical (e.g., manipulate and move objects around) and in cognitive-behavioral ways (e.g, verbal feedback from virtual humans). At user level, there are the learner's individual factors (e.g., ways to engage and to stay alert and interested in the simulation, identification with a virtual self who has a role in the narrative). These help them control the events and transfer the learned skills in a real-world situation. At content level, there is the narrative, with the story where users take control and can experience the environment first hand. The interpersonal, social, and cultural context are the last of the proposed presence factors. Here is where social presence is mostly

visible, as F. Mantovani and Castelnuovo (2003) argue that creating social interaction in a learning environment is at times more important than image realism of the story background, especially if the characters are anthropomorphic. The authors close with the social and cultural context which underlines the importance of ecological validity of the virtual content. In this sense, the cultural meaning and the social context ought to be reflected and be relevant to the user, in order to enhance the sense of presence.

One element that appears more or less explicit in each of the above factors is interaction. Initially discussed under perceptual factors, it becomes evident that a sort of interaction and feedback from the virtual characters is a necessary element to fulfill the role of each of these factors in fostering presence. For instance, training social skills may require a narrative where people interact verbally and nonverbally with virtual characters, and if these are not able to fulfill these tasks the recreation of an environment where actors function in an ecological way is hindered. It can be therefore hypothesized that a model of presence that fits the needs of virtual training with intelligent characters might require a separate factor dedicated to multimodal interaction.

The review of definitions has shown so far the complexity of presence compared to the more straightforward immersion concept. In spite of their complexity differences, they are closely connected, with immersion levels impacting on the experienced presence (Slater & Wilbur, 1997). Cummings, Bailenson, and Fidler (2015) conducted a meta-analysis to identify the effect of immersion on presence levels reported by users. The authors refer solely to spatial presence and argue that studies on self-presence or social presence were insufficient for such an analysis. The analysis shows that immersion has an overall medium effect on presence ($r = .316$). At a closer look, image proved to have a small effect similar to that of sound, whereas technical features such as stereoscopy, field of view, and tracking level proved to be more important in determining spatial presence (Cummings et al., 2015). The authors argue that these findings should be considered with care, because they reflect only spatial presence and designers might have to pay attention to other technical aspects (e.g., multimodality and communication synchronicity) when addressing self and social presence in immersive environments.

2.2.4 The role of immersion and presence in virtual reality applications

One pervasive questions in VR training application research has been whether the qualities of being immersive and of fostering presence make applications effective in reaching their purpose (Botella et al., 2010; Price & Anderson, 2007; Tortella-Feliu et al., 2011). This can have major implications for planning, design, and cost allocation. Whereas some training and treatment practices require a certain level of immersion and presence, there is no proof that these ought to be constant throughout all applications across the field. Therefore, depending on the purpose of the application, one could assign appropriate levels of immersion and presence and implement the most fitting technology.

Several studies detailed below show that achieving both immersion and presence proves important in skill training and phobia treatment in virtual environments.

Levels of immersion are usually discussed in relation to CAVE, CAVE-like environments, HMDs, and desktop computer, in various comparative studies. There is earlier evidence that increased immersion (an L-shaped projection display) has a positive impact on task performance compared to lower immersion (laptop screen) for tasks where visual memory is relevant (Wang & Bowman, 2007). The authors suggest that further studies ought to clarify which elements of immersion have this effect, and which components among field of view, screen size, or field of regard are at work. (Ruddle, Payne, & Jones, 1999) show that HMDs helps with faster navigation in the virtual environment compared to a desktop display. An informal comparison between task performance (military checkpoint routine) in CAVE and in a desktop simulation conducted by Loftin et al. (2003) suggests that people perform better the immersive environment.

Immersive environments successfully help patients overcome fear of heights, compared to a control group (Krijn, Emmelkamp, Biemond, et al., 2004). However, for fear of flying, Tortella-Feliu et al. (2011) demonstrate that immersive VR treatment is equally effective as the training on a desktop computer, both at posttreatment and at follow-up one year later, suggesting that more common devices than the CAVE or CAVE-like environments can be implemented to treat at least certain

phobias. The same findings are supported by a study on internet-based treatment for fear of public speaking, which was successful in lowering anxiety levels with the help of an online interface (Botella et al., 2010). In comparing HMDs with desktop applications, Robillard, Bouchard, Fourier, and Renaud (2003) report that less immersive environments manage to induce anxiety in phobic subjects and suggest, like Botella et al. (2010) above, that it is possible to use more affordable environments (i.e., desktop applications) in treatment.

Apart from image quality, sound is an affordance that belongs to immersive characteristics and it can be used to foster feelings of presence (Hendrix and Barfield, as cited in Västfjäll (2003)). In the case of fear of flight for instance, head-mounted display simulations contain the plane engine sounds during takeoff and landing, flight attendants talking, as well as simulated weather effects while in flight (Rothbaum, Hodges, Smith, Lee, & Price, 2000). In posttraumatic stress disorder (PTSD) treatment with a head-mounted display, sound helps patients recall anxiety-provoking situations and practice desensitizing situations (Difede & Hoffman, 2002; Ready, Pollack, Rothbaum, & Alarcon, 2006). In attention deficit disorder (ADD) treatment, ambient classroom sounds are also used to understand, assess, and help rehabilitate attention processes (Glanz, Rizzo, & Graap, 2003)

With presence, the discussion moves away from visual or aural fidelity into the realm of emotion elicitation. The correlation between presence and emotion though is debated across various studies. Slater suggests that presence and emotion are orthogonal, which means that there should be no significant relation between them, as one can feel present yet uninvolved emotionally (Slater, 2003). However, several empirical studies contradict this statement. Presence has been studied in relation to emotion and (Baños et al., 2008) showed that, independent of mono or stereoscopic view (desktop versus immersive display), people who experienced more presence reported also stronger positive emotional reactions. However, the reported findings are only correlational and researchers argue that either experienced emotion leads to presence or the other way around.

In a study on effects of virtual exposure therapy on acrophobia, subjects reported experiencing more presence in the CAVE compared to HMDs, yet higher presence didn't lead to a more effective treatment for either of the immersed groups (Krijn, Emmelkamp, Biemond, et al.,

2004). Moreover, researchers found no correlation between presence and anxiety during the exposure sessions. In a later study by (Price & Anderson, 2007), presence and anxiety were found to be correlated, and researchers have further sought to identify the direction of the correlation. According to the study results, the more present people declared to feel in the virtual environment (HMDs), the higher the measured anxiety. Researchers suggest that presence acts like an enabler for emotions (anxiety in this case) during exposure, and that people with greater phobic anxiety might experience more presence and therefore more anxiety in the immersive virtual environment. However, experienced presence levels had no direct influence on the treatment outcome, and researchers explain that presence is necessary to elicit anxiety during exposure, but it might be insufficient by itself to extinguish the fear. The study of (Bouchard, St-Jacques, Robillard, & Renaud, 2008) reports otherwise that state anxiety impacts directly the experienced presence in immersive environments, but only when measured with brief verbal accounts during the experiment and not on overall presence measured with the Witmer and Singer (1998) questionnaire. They further indicate that this raises questions about the instrument itself. Researchers reported that presence measured with the questionnaire actually dropped in the anxiety producing condition, which they explain through the fact that the questionnaire rather measures variables that influence presence than the subjective experience of presence itself.

Robillard and colleagues compare as well the relationship between presence and anxiety in phobic and non-phobic subjects (Robillard et al., 2003). They report that a high experienced anxiety correlates significantly with a high reported feeling of presence. However, the direction of this correlation is further discussed but not explored empirically. Moreover, a less immersive environments proved anxiety-inducing for the phobic subjects as well. An interesting finding of the study is that non-phobic subjects experienced less anxiety and less presence as well.

Immersion and presence are often seen in a causal relationship, and the meta-analysis of Cummings et al. (2015) brings into discussion the impact of communication channels, multimodality, synchronous and asynchronous communication, as well as interaction with artificial intelligence (AI) on social and self-presence. In this respect, the presence of artificial intelligence and interaction with the virtual

characters helps fulfill various tasks in virtual environments. This is particularly visible in medical training, where the implementation of virtual humans has educational advantages. One such case is Just VR developed by Manganas et al. (2004) that recreates emergency situations with a virtual patient and a virtual assistant. Here, a trainee is faced with a crisis situation where they have to guide the virtual assistant in helping the patient. A small pilot trial showed that trainees reported being convinced of the reality of the simulation and the authors suggest that this tool can train medical personnel with decision making under time pressure.

In other studies, although not explicitly measured, virtual patients elicit the feeling of sharing a common virtual setting with another person, reflected in trainees' behaviors towards standardized virtual patients (e.g., empathic behaviors and signs of rapport as seen in the study of Raij et al. (2007)). In the comparative study between a real and a virtual patient (2D interactive projection on a wall), students managed to conduct a patient interview in a similar manner (ask the same amount of critical questions) both with a real and a virtual patient, in spite of the latter's limited verbal and nonverbal expressiveness (Raij et al., 2007). They rated the virtual environment equally educational as a real one and manifested similar empathy (although rated as less genuine) as in a real life setting.

Attention to realistic representation and behavior is visible also in the design study of Rizzo, Kenny, and Parsons (2011). Here the authors create Justina, a virtual female patient with PTSD caused by sexual assault. Trainees are required to conduct clinical interviews with her in a desktop-based setting. Trainees rated the system as above average believable. Moreover, they also reported being interested to work with such systems to improve clinical interview skills in the future (Rizzo et al., 2011).

The studies reviewed above represent only a part of the large body of work on virtual environments and their positive impact in training and treatment. Both immersion and presence are complex core constructs of the VR research field and they ought to be carefully considered in the planning of treatment and training procedures. Whereas immersion is well understood in the literature, presence appears less clear cut. Moreover, as F. Mantovani and Castelnuovo (2003) suggest, presence alone might not be sufficient during training, and it might actually be

good for trainees to pause and reflect (break the presence) on what is happening. Baños et al. (2008) state that one might actually not even need sophisticated, non-stereoscopic technologies to elicit emotional responses, as long as people experience high levels of presence with simpler ones like desktops.

However, this might not be the case for all types of training applications. For public speaking training purposes, immersive environments might be the technology of choice due to the fact that they allow a realistic representation of the complex speech setting. Speaking to a crowd is an experience that implies many setting items (e.g., a podium, a stage, lightning, room echo), speech-specific procedures (e.g., entering a hall, waiting for a hall to fill up with audience members), and relevant cues from the audience (e.g., facial expressions, body movements, localized sounds such as conversations or comments). Due to its size and its immersive features (e.g., stereoscopic images in a cube of projection screens (Cruz-Neira et al., 1993), spatial sound (Poeschl, Wall, & Doering, 2013)), a CAVE could reproduce many of the speech setting details at a realistic scale, compared to a desktop application. At this point one could only hypothesize that such details are indeed needed for the success of public speaking training, but in order to accept or reject this hypothesis, CAVEs should be implemented and studied in public speaking studies. It follows that, before dismissing immersive virtual environments on the account that desktop applications are sufficient in public speaking phobia treatment, one should first test their success in skill acquisition with speakers. Further work should be dedicated to this endeavor as well as to comparative studies between desktop applications and immersive environments for this specific type of training, in order to understand how much immersion and presence are necessary for maximizing speech training efficiency with minimum design effort.

2.2.5   Speech anxiety treatment in virtual reality

2.2.5.1   What is fear of public speaking?

Fear of public speaking is a type of social phobia that involves having to talk in public and fearing the scrutiny and evaluation of others while doing so (Heimberg, Hope, Dodge, & Becker, 1990). Speech anxiety was identified as the most common manifestation of social anxiety. It is followed by entering a room occupied by others, meeting with strangers, eating in public, and writing in public (Faravelli et al., 2000). With regard to its integration within the social phobia spectrum, speech anxiety falls under the performance type subcategory, which describes activities performed comfortably only while being alone (other examples include eating, writing, and urinating in public) (Heimberg, Holt, Schneier, Spitzer, & Liebowitz, 1993). Kessler, Stein, and Berglund (1998) found that social phobia with pure public speaking fear is more prevalent among those with high school education, and less in collage graduates. The authors argue that people with less education than high school might have lower chances of having their fear activated, since it is rare for them to speak in public.

Speech anxiety manifests through psychological symptoms (e.g., negative self-evaluation), physiological symptoms (e.g., elevated heart rate, intensive breathing, higher adrenaline and noradrenaline secretion, intense sweating, etc.), and motor-behavioral symptoms (e.g., limited movements, loss of control of movements such as awkard gestures, unusually high voice pitch, interruption of line of thoughts and speech, etc.) (Beushausen, as cited in Koholka (2012)). Internally, people constantly monitor the feelings that transpire to the outside such as hot feelings leading to blushing or sweating, whereas externally they tend to distort negatively the verbal and nonverbal feedback received from audiences (Rapee & Heimberg, 1997). Anxious people focus more on themselves and less on environment (e.g., audience), they make more nervous movements, and are also evaluated less favorably by audiences than low anxious people (Daly, Vangelisti, & Lawrence, 1989).

In the communication field, authors use the term "communication apprehension", which refers mainly to the anxiety someone feels in relation to either a real or anticipated communication with other person or persons (McCroskey, 1977). Miller and Stone (2009) group under oral communication apprehension several fear types which

include, along with group discussions, or meetings, also public speaking. Glossophobia, or speech anxiety represents therefore the clinically diagnosable fear of public speaking (Miller & Stone, 2009). Since communication apprehension has not been researched in relation to virtual reality, the terminology adopted in the following sections borrows from the clinical research on phobia and speech anxiety.

2.2.5.2   Speech anxiety treatment

Several treatment methods are available and they can be grouped into three therapy types, based on a meta-analysis of self-report effectiveness of public speaking treatment techniques (Allen et al., 1989):

- Systematic desensitization changes the associations a person makes with the speaking situation. It makes the person face the feared stimuli gradually until more pleasant responses are associated with the speaking situation. Through several treatment sessions, the therapist guides the patient gradually through the speech situations they fear, from the least feared to the most feared ones, and teaches them relaxation skills. When all situations are covered and patients feels no anxiety anymore, the treatment is considered successful.

- Cognitive modification targets the beliefs of a person with regard to the feared situation. The therapist discusses with the patient about the feared situations and points out irrational thoughts the patients have in those situations. This method helps patients recognize such irrational thoughts and manage to diffuse them. These irrational thoughts (e.g., people always make fun of me) are gradually replaced with rational ones (people laugh only if they hear a joke, otherwise they pay attention), and the speaker is taught to focus on the latter ones during a speech.

- Skill training starts with the premise that people feel insecure because they lack certain speaking skills. Such skills can be learned, and mastering them helps raise the confidence of the speakers. Such skills are learned through intensive speech practices.

One of the most effective treatments against fear of public speaking belongs to the cognitive modification type, due to its long lasting results (Fava et al., as cited in P. L. Anderson et al. (2005)). Cognitive

behavioral treatment (CBT) consists of a practice that teaches patients how to adapt to their inter and intrapersonal worlds through cognitive and behavioral learned skills (Heimberg, 2002). CBT is usually coupled with two types of exposure: imaginal and *in vivo* exposure. The first requires a patient to imagine the feared situation, the second exposes the patient to real fear stimuli (Bush, 2008).

Exposure is central in most CBT approaches and through this technique patients have to face the feared situation while staying psychologically engaged in order for the habituation and extinction of fear processes to occur (Heimberg, 2002). Through exposure, patients experience a reduction in anxiety (habituation) and the anxiety-provoking event they anticipate to happen following their exposure becomes absent (extinction) (Wallach et al., 2009).

2.2.5.3 Virtual reality exposure treatment

In spite of its importance in the CBT practice, exposure to feared situations can be very hard to stage in the case of several phobias. In treating fear of flights, the subject would have to actually buy a plane ticket and take a flight; in the case of agoraphobia, the person would actually have go outside and risk the embarrassment of meeting known people during the treatment session. Virtual reality offers an elegant alternative to the exposure component of phobia treatment and several feared situations can be simulated in the safe environment of a laboratory.

Virtual reality exposure therapy represents a treatment procedure where the subject faces a computer-generated simulation in order to be exposed to specific fear-inducing stimuli (Parsons & Rizzo, 2008). VRET proved to be a viable alternative to both imaginal and *in vivo* exposure. The first one has low costs and confidentiality, but has low success rates; the second one has high success rates, but is expensive and has low therapist control (Bush, 2008; Wiederhold et al., 2002). Through VRET, patients get to experience anxiety-eliciting situations similar to *in vivo* ones while still remaining in a safe environment (Wiederhold et al., 2002). Still, in order to be an effective tool, it needs to meet three conditions:

- Patients need to feel present in the simulated environment, hence be part of the simulation rather than watch it passively (Slater,

Pertaub, and Steed, as cited in Krijn, Emmelkamp, Olafsson, and Biemond (2004)).

- The virtual environment triggers the anxiety that needs to be treated (Hodges et al., as cited in Krijn, Emmelkamp, Olafsson, and Biemond (2004)).

- Extinction of anxiety and cognitive changes following the treatment are transferred to real-life situations (Krijn, Emmelkamp, Olafsson, & Biemond, 2004).

Virtual reality exposure therapy has been already applied for panic disorder (Botella, Villa, García-Palacios, Baños, & Perpiñá, 2004), acrophobia (Emmelkamp et al., 2002), arachnophobia (Garcia-Palacios et al., 2002) or PTSD treatment (Rizzo, 2008). Speech anxiety is part of this trend and has also been successfully explored with the help of this technology (P. L. Anderson et al., 2005; Pertaub, Slater, & Barker, 2001; Safir et al., 2012).

2.2.6   Virtual humans in speech anxiety treatment applications

2.2.6.1   General characteristics of virtual humans

The main component that facilitates the experience of anxiety in the virtual reality speech anxiety treatment scenarios is the virtual audience. A virtual audience is composed of virtual characters, which populate a simulation (Magnenat-Thalmann & Thalmann, 2004) and bear various names throughout the literature. There are the avatars that embody a real user in the virtual environment, but cannot act in the absence of a user's control (Cassell & Vilhjálmsson, 1999; Fabri, Moore, & Hobbs, 1999). By contrast, the Embodied Conversational Agents (ECAs) have high cognitive and behavioral capabilities and can not only understand and produce speech, but also respond with natural-like behaviors, through facial expressions or gestures (Cassell, 2001; Kasap & Magnenat-Thalmann, 2007). Figure 2.3 depicts an ECA used in a study on gender difference in impression management (Derrick & Ligon, 2014).
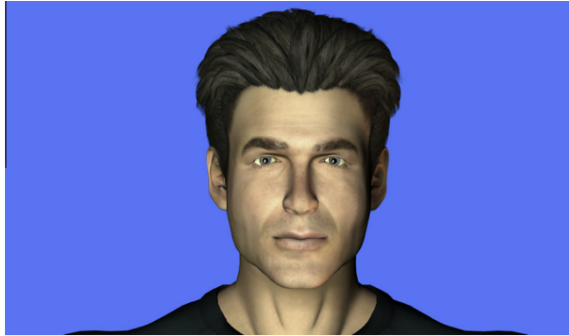
Figure 2.3. Impression management embodied conversational agent
(Derrick & Ligon, 2014, p. 43)

Virtual patients used in medical interview skills training belong to this
category as they are able of interaction with trainees. Nunamaker,
Derrick, Elkins, Burgoon, and Patton (2011) propose a different
version of an embodied conversational agent called the Special Purpose
Embodied Conversational Intelligence with Environmental Sensors
agent (SPECIES). This is an automated kiosk that represents an
embodied conversational agent which serves with the port-of-entry
screening process (Nunamaker et al., 2011). All information is acquired
via multiple sensors, such as cameras and eye trackers (Nunamaker
et al., 2011). The virtual audience concept proposed here follows the
embodied conversational agent's ideal characteristics, therefore the next
sections will present virtual humans with high degree of autonomy.

The three levels of autonomy proposed by Kasap and Magnenat-
Thalmann (2007) are *perception*, *decision making*, and *action control*:

1. At perception level, an autonomous virtual character perceives
   its environment in a realistic manner, as a human would do.
   Through visual, aural, and sensory perception virtual humans can
   navigate their environment freely. Within a virtual audience, this
   would translate into characters whose behavior would be shaped
   by the spaces they populate and by the actions of other virtual
   characters they encounter. One can call this a group dynamic
   feature because it puts the virtual human within the context of
   a group of other virtual characters, each with its own autonomy
   and perception.

2. The decision making level ensures that the character has the ability to evaluate a situation and choose the proper action in response. These processes rest on intelligence, motivations, plans, etc. A virtual audience is faced with a double challenge. One the one hand, just as seen above, virtual humans in a group ought to perceive their environment and the actions of fellow characters as well as those of the speaker. On the other hand, they should also take decision based on the mitigation of multiple sensory inputs that may occur simultaneously (e.g., several audience members talking at the same time).

3. At the action control level, autonomous agents perform movements that denote their interactive capacity (e.g., emotional reactions, gestures in conversations). Other types of movements mentioned by Kasap and Magnenat-Thalmann (2007) are functional movements that relocate the body, such as walking or grabbing objects. Both these kinds of movements are relevant for a virtual audience: the first type signifies understanding of and interactivity with the environment (e.g., with the speaker or other virtual humans); the second type ensures that characters act upon goals (e.g., enter the room and sit, pick up a pen and start taking notes).

The authors further distinguish between interaction and personification, as two other features of autonomous agents. Interaction contains nonverbal and verbal behaviors used by the virtual agent to communicate with the user, such as facial expressions, gestures, and dialogue (Kasap & Magnenat-Thalmann, 2007).

The last feature–personification–refers to assigning personality and emotion characteristics to virtual agents (VA) (Kasap & Magnenat-Thalmann, 2007). The VA's personality can be a relevant component in one-to-one interaction training, where it can mediate the interaction between the two actors or be used as diagnosis tool (Krishnan, Foster, Kopper, & Lok, 2012). One way to simulate personality is to apply the five-factor model of personality proposed in the psychology field (Kshirsagar & Magnenat-Thalmann, 2002). For realistic virtual audiences, personality may have less relevance. As part of a group of listeners, people may not show many personality features towards a speaker in the absence of closer encounters. Emotions on the other hand are relevant because they are easy to spot in facial expressions

and they can serve as feedback in an interaction (e.g., approach or
avoidance response). Among the models used for emotion simulation
are the OCC model (Kiselev, Hacker, Wankerl, Abdikeev, & Nishida,
2010) and appraisal models (Wehrle, Kaiser, Schmidt, & Scherer, 2000).

2.2.6.2 Virtual audience characteristics in speech anxiety treatment

P. Kenny et al. (2007) suggest that successful and engaging virtual
agents possess qualities such as believability, responsiveness, and
interpretation skills. They ought to be believable so that they help
the user play along the depicted scenario. They have to be responsive
both to users and to their own virtual environment and they have to do
that using the same verbal and nonverbal codes as a real person would
in that given situation. Translated to a virtual audience, the characters
would have to be internally able to respond to a dynamic environment
and to do this in a realistic way (P. Kenny et al., 2007).

In spite of intelligent ECAs who can make conversations with users
and respond nonverbally in appropriate manner, virtual audiences
depicted in many studies have little autonomy to interact with a
speaker, as the literature review below will reveal.   Most of the
audiences represent treatment tools in speech anxiety or social phobia
treatment.   The studies on crowd movement simulation represent
few exceptions, where virtual-virtual human interaction, movement
and avoidance models, and psychological models of behaviors were
studied (Silverman, Pelechano, & O'Brien, 2005; Ulicny & Thalmann,
2002). An explanation for lack of autonomy could be that, in phobia
treatment, therapists hold control over the procedure, so that exposure
occurs gradually and is performed in a safe environment.   Another
explanation could be that in phobia treatment the accent is put on how
users gradually adapt to a given scenario, through repeated exposure,
and not on how the scenario adapts in real time to the user (North,
North, & Coble, 1998).   The following literature review introduces
several speech anxiety studies in virtual environments and reviews a
few characteristics of virtual humans relevant for the virtual audience
concept proposed in the dissertation.

In one of the first studies with virtual humans in anxiety treatment
simulations, North et al. (1998) used a virtual auditorium with up
to 100 virtual humans.   These could give prerecorded verbal and
nonverbal feedback to the speakers, such as laughing, encouraging the

speaker, or ignoring them. The way audience feedback was created is not described in the article. A study by Slater, Pertaub, and Steed (1999) used a male-only virtual audience that could follow the speaker move around with the eye gaze and head turns. Other behaviors included yawning, sleeping, and the facial expressions for the six basic expressions (happiness, contempt, sadness, anger, fear, and surprise). Moreover, VHs could interrupt the speaker or even leave the room. The study reports that the behaviors were manipulated by a person who could listen to the speech and trigger the behaviors in accordance with the content of the speech. Two audience types were used: a very friendly and a very hostile one, in an immersive and a computer display visualization condition. The same method of audience behavior manipulation was used in the study by Harris et al. (2002). Therapists manipulated the scenario by gradually filling up a virtual room with people. They could also select applause, laughter, or conversations among audience members. In the study of P. L. Anderson, Rothbaum, and Hodges (2003), virtual people could be controlled to display boredom, interested looks, applauding or neutral behaviors. Grillon, Riquier, Herbelin, and Thalmann (2006) proposed several scenarios for social phobia treatment, among which there was also an auditory filled with male and female virtual humans who were equipped with prerecorded sentences and facial animations to match the speech. Therapists would trigger these behaviors during exposure sessions. In both the initial study of Wallach et al. (2009) as well as in the follow-up study (Safir et al., 2012) the used virtual audiences was offered by Virtually Better, a company specialized in virtual reality simulations. Audiences were not described, but researchers mentioned that their behaviors were controlled by the therapist.

Whereas these studies used animated virtual humans, other studies used video recordings of various behaviors played by people who are then integrated into the virtual environment. Therapists would again control which behaviors are displayed (interest, boredom, applause), in accordance to the therapy program (P. L. Anderson et al., 2005; J. M. Lee et al., 2002).

One can use virtual audiences that are either created and animated with graphic tools or that are made of videotaped people. Both can be controlled from outside, by choosing which behaviors should be shown to the speaker. Still, only the animated one can be autonomous in

Table 2.1
Virtual audience typology - own depiction

	Source	
External control	Animated	Recorded
Autonomy	Proposed audience for public speaking training	-
Display technology	Computer display, HMD, CAVE	Computer display, HMD, CAVE

the sense described above. Table 2.1 summarizes display methods and autonomy levels for state of the art virtual audiences.

The concept for the training audience proposed in the next chapter takes as starting point the autonomy levels discussed so far and also the features of existing virtual audiences reviewed above.

2.2.7 Summary

The chapter reviewed virtual reality technology, its applications in various fields, and relevant components, such as immersion, presence, and virtual humans. Virtual reality is an accessible alternative to training in many fields (e.g., medical procedure training, patient diagnosis, flight or military training) as well as in phobia treatment where patients are treated in a safe environment. Complex simulations, such as those that recreate human interaction, require architectures that support virtual humans with high degrees of autonomy. These can understand speech, react to users' input, talk back, and enact their own nonverbal behaviors.

Using such VHs is becoming common practice in fields that require one-to-one communication skills, such as patient diagnosis or interviews. In different settings though, where communication is from one to many, such as in presentation and public speaking situations, virtual reality has been used far less. Virtual audiences have been typically

implemented for speech anxiety or social phobia treatment, and their characteristics have yet to be debated in relation to other applications. The literature overview showed that virtual audiences have little autonomy and are usually controlled by a therapist. Moreover, the verbal and nonverbal behaviors that they enact are rarely described in detail. For this reason, a new audience type is proposed, one that ought to: (1) have more behavior autonomy than existing models, (2) react in real time to the speaker, and (3) depict behaviors in accordance with those displayed by audiences in real speech situation. The next chapter presents the concept of this audience and sheds light on its needed characteristics in the field of communication skills training, away from the therapy office.

CHAPTER 2.3

Communication training in virtual reality: Audience research model proposal

The concept for a virtual audience in virtual reality proposed here rests on two major fields discussed in the previous chapters: communication skills training and virtual reality. From the communication skills training perspective, there is a great variety in opinions on what exactly communication skills should contain. Basic skills address types of messages and audience adaptation, whereas advanced ones address heavy reasoning skills, attitudes, and audience analysis (Morreale & Backlund, 2002). However, opinions converge in the point where communication skills require a solid knowledge of various types of presentations (e.g., informative speaking, persuasive speaking) and of delivery styles (e.g., extemporaneous and impromptu delivery), as well as other skills such as organizing the body of speech, handling question and answer sessions, and listening (Johnson & Szczupakiewicz, 1987). The communication in the disciplines and communication across the curriculum perspectives pointed out that students should be able to express themselves appropriately both in their field of expertise, but also outside of it, for instance in front of clients or experts in other fields (Dannels, 2001, 2002; Dannels & Housley Gaffney, 2009). This requires not only mastering speech skills, but also adaptability to various audiences (Wiederhold & Bouchard, 2014).

Virtual environments are a solid platform that has supported various training and treatment applications in the past 20 years and virtual audiences are one of the success stories in speech anxiety treatment. In order to move speech and presentation skills training into the virtual environments, the existing types of virtual audiences require adaptation at different levels. Just like in the case of anxiety treatment where virtual audiences match the desensitization steps of patients, a virtual audience for public speaking skills would have to adapt to the speech practice steps of trainees.

When communication training and speech anxiety are put next to each other, one can notice two major differences between treatment and training:

- Research focus of attention: in speech anxiety treatment, the focus of research is on how to apply stimuli to elicit and lower anxiety levels (Wallach et al., 2009, 2011). Such stimuli are various audience types which can enact positive and negative behaviors. In anxiety treatment, improvement is measured through measuring anxiety levels (McCroskey, 1970). In communication skills training, competence and improvement are assessed through performance measures such as The Competent Speaker Evaluation Form (Morreale et al., 2007; Perotti & DeWine, 1987).

- Training/treatment components: from a procedural perspective, both patients and trainees would have to give a presentation in front of an audience. Both may feel anxious and habituation with the public speaking situation can be beneficial for both. However, whereas the patient finishes treatment once anxiety levels are lowered (such as in Wallach et al. (2011)), the trainee might continue exposure as long as needed to improve presentation skills and speech content to various speech situations. Presentation content is rarely a subject of discussion in speech anxiety treatment studies, whereas it represents one of the most important elements of communication training, as it will be seen in the expert interviews conducted in the first study of the dissertation.

These differences reflect in the way the virtual audience could be adapted for skill acquisition purposes. Therefore, having a virtual audience with expanded functionality towards interactivity and customization is a desideratum that turns it into a key driver of skill improvement.

2.3.1   Audience centrality

As Verderber et al. (2008) point out, effective speakers are audience-centered. Such speakers offer ideas which are relevant for the listeners and have rhetorical appeal. The content is relevant to the listeners, is presented in a structured manner, and is delivered expressively. The authors propose a model of speech effectiveness that contains three major elements: speaker, audience, and speaking context. There are
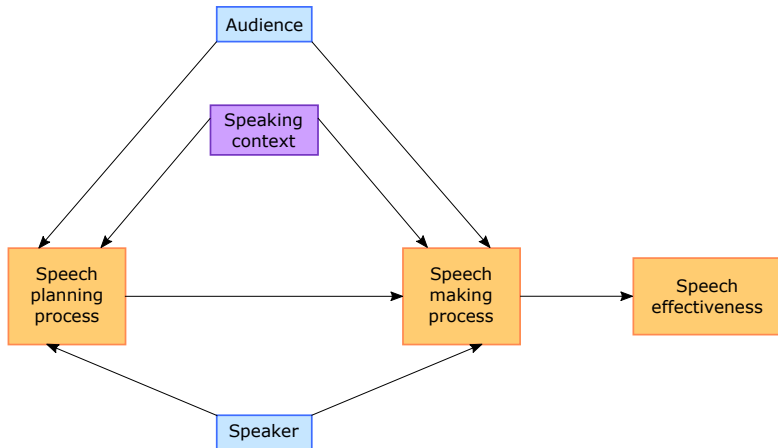
Figure 2.4. The public speaking effectiveness process model
(Verderber et al., 2008, p. 4)

two processes which unite these elements: speech planning and speech
making. The outcome of the model is speech effectiveness. This model
depicted in Figure 2.4 underlines the centrality of audiences.

The speaking context comprises a physical speech settings, a cultural
one, a historical, and a psychological setting. They all relate to the
audience: how far or close a speaker stands from the audience, what
are the values and believes of audience members, what events prior to
the speech might have affected them, in what mood they are. These
elements can impact on how a speech is perceived.

Two processes unite speakers, audiences, and context. The first one
takes place before the speech, the second one during the discourse.
Speech planning refers to the steps a speaker takes to organize the
information and the delivery style in a speech (Verderber et al., 2008).
The focus is primarily on content preparation (what's the goal of the
communication, how to gather relevant information, etc.). The speech
making process represents the actual speech delivery phase, and, as
the authors say, is the part where the speaker focuses on both the
delivery and on the feedback they receive from the audience during the
presentation.

Speech effectiveness represents the goal of public speaking and presentation training programs. Speech effectiveness is defined as "the extent to which audience members listen to, understand, remember, and are motivated to act on what a speaker has said" (Verderber et al., 2008, p. 7). A speech is effective as long as it is audience-centered, it delivers audience-appropriate content, is well-structured, and is delivered enthusiastically Verderber et al. (2008). These characteristics can be empirically studied in the future in order to measure and quantify the impact of each characteristic on speech effectiveness. Nevertheless, the model and the element definitions underline clearly that public speaking is a skill where content has to be properly developed, tailored, and delivered to the right audience.

2.3.2   Research model for a virtual audience

Mirroring the model described above, the virtual audience research model synthesizes what elements ought to be studied in order to create a virtual audience design concept that can prepare speakers to become effective performers. At the center of this research model lies the virtual audience. The training setting brings the actors together: the speaker (in black), the audience (in green), and the trainer (in red). Their interactions (marked in purple lines) are mediated by sensors (in yellow) and by the graphic user interfaces (GUI in yellow). These actors and their interactions are subsumed to the communication training context (marked in violet). One can distinguish three levels at which the audience concept can be described and studied, depicted in Figure 2.5.

1. The first level addresses the virtual audience (in green) and its technical and human-like attributes (in black).

2. The second level describes the virtual reality context where interactions take place. On the one hand, there is the interaction between speaker (trainee) and virtual audience, which is marked by purple two-way arrows. The speaker gives verbal and non-verbal input and these can be interpreted by the virtual audience with the help of sensors. Based on their interpretation, the virtual listeners respond accordingly (the purple arrow from audience to speaker). On the other hand, there is the interaction between trainer and virtual audience (as counterpart to therapist audience control in anxiety treatment). The trainer can control
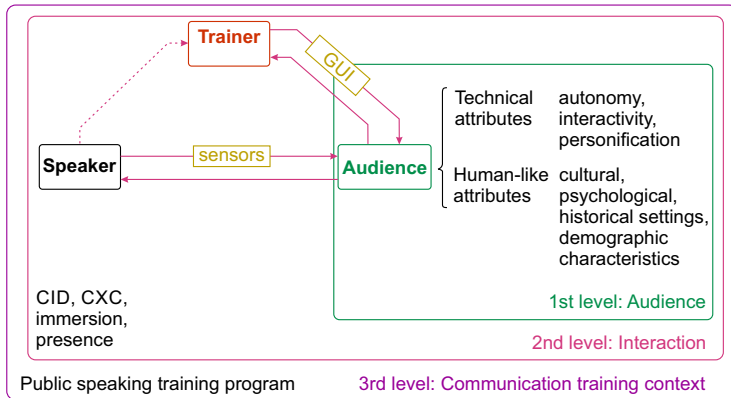
Figure 2.5. Proposed research model for a virtual audience

through a GUI (in yellow) the way audiences respond to a speech (the purple arrow from trainer to audience). Furthermore, the trainer may as well observe the audience behavior display (the purple line from audience to trainer), as shown in the immersive virtual environment. The dotted line between speaker and trainer represents the fact that the trainer can also monitor the speaker and base its audience control decisions on the speaker's performance.

3. The third level of the model represents the broad public speaking and presentation training program.

Next each level is described in detail in light of the communication and VR research.

2.3.2.1   First level

The first level is dedicated the the audience alone. It contains two parts: one for the artificial intelligent elements of autonomous agents (Kasap & Magnenat-Thalmann, 2007) (technical attributes) and one that stands for potential human-like traits (human-like attributes) (Verderber et al., 2008). Within the group of virtual humans, audience members should be autonomous enough to navigate the virtual environment they inhabit in a coherent manner, be interactive with and respond to the actions of fellow artificial characters, as well as react as a group (Forsyth, 2009; Kasap & Magnenat-Thalmann,

2007). Moreover, they ought to embody personality and emotions (personification) that would justify their actions (Kasap & Magnenat-Thalmann, 2007).

The part that stands for the human-like characteristics entails their cultural, psychological, and historical setting, as well as their demographic characteristics (Verderber et al., 2008). Ideally, a virtual audience would have a cultural, historical, and psychological background, similar to real audiences. However, design constraints ought to be carefully considered in relation to this level of virtual audience background, in order to make it viable and easy to use.

2.3.2.2   Second level

This level contains the interaction between speaker and audience, as well as the one between a trainer and the audience. Moreover, it proposes two methods to control the virtual audience as well as enable interaction: a graphic user interface (marked as GUI) and sensors.

The speaker gives away verbal and nonverbal cues and the audience performs its feedback as well. Depending on the degree of autonomy of virtual characters, this feedback can be genuine and reflect a realistic understanding of speech or can be simulated and in reality be controlled by the trainer. Various sensors (e.g., eye tracking, microphones) would help pick up speech content and delivery style (e.g., gaze direction, voice volume, body movements) and feed them into the system to process an appropriate audience response. If this is not implemented, trainers would control audience reactions, with the help of a GUI, like therapists do in anxiety treatment scenarios (e.g., Grillon et al. (2006)). Interaction between speaker and audience is explored at large in the first study of the dissertation. Suggestions are also made with regard to the graphic user interface.

Two further components of the second level are the communication in the discipline (marked as CID) and communication across curricula (marked as CXC) paradigms, which establish the context of the speech in real life: whom is the speech going to be addressed to? The answers to this question can further help establish audience characteristics, speech content, and auxiliary elements, such as props (e.g., presentation slides). These paradigms have not been specifically inquired in the studies reported here, but can guide future research once the application is running.

Immersion and presence can be included in this level as well, because they occur through the interaction between speaker, immersive technology, virtual humans, and the applications they inhabit. The virtual audience would be displayed in a highly immersive CAVE-like environment that surrounds the user with 3D images and sounds, while shutting out other sensorial inputs. Moreover, as long as immersion is high and the virtual audience acts realistically, speakers might also experience presence. Since both immersion and presence were related to successful skill transfer (F. Mantovani & Castelnuovo, 2003), they are also desiderata to speech training success. They represent the goals to achieve through the design of virtual audiences and are subject of future studies with public speaking trainees.

2.3.2.3   Third level

The last level encompasses the context of training with a virtual audience. Speech practice is only a part of the speech and presentation skills training program, where various skills are first taught in theory, before moving on to practicing them (Beebe, 2007). Sánchez, Barreiro, and Maojo (2000) speak of a similar context–the real environment– where education takes place (e.g., a chemist's laboratory or a pilot's flight simulator). The authors indicate that this training context is the base for choosing the source knowledge such as concepts, skills, or information related to the subject to be learned. Moreover, if virtual audiences become largely available for speech training sessions, future studies could help establish the best ways to use them from a financial and organizational perspective, relative to the whole public speaking training seminar. The role of speech practice was not inquired specifically in the interviews, however it emerged inductively and therefore was further examined in the training seminar observational study.

Borrowing from the speech effectiveness model presented in Figure 2.4, it follows that a successful training with a virtual audience would lead to speech effectiveness outside the training setting. Such assumptions could be verified empirically in future work. Moreover, since the virtual audience presented here is suggested for immersive environments, its implementation in commercially available platforms can also become subject of research.

2.3.3 Summary

The public speaking effectiveness process model by Verderber et al. (2008) offered a background for the creation of the virtual audience research model presented in Figure 2.5. This diagram establishes the main points of research in the field and puts them into the broad context of speech effectiveness in real life situations. For instance, one can start the research from the outer level towards the inner ones and search for what makes training effective and how does each level impact effectiveness. One can also look the other way around and propose a virtual audience and see how it can foster immersion and presence. One can search for ways to measure effectiveness of training in virtual reality or challenge the whole effectiveness concept completely and find further outcomes of such training.

The audience conceptualized here is exemplified in the last chapter of the dissertation, through a virtual audience design concept and a concrete scenario. The virtual audience design concept suggests various speech settings, and the scenario exemplifies several nonverbal behaviors the audience could display, in a time frame of five minutes. All three levels reflect in the findings of the empirical studies of the dissertation. The first study which will be presented next tackles interaction and audience characteristics, such as familiarity with the speaker. The second study focuses on listeners' nonverbal behaviors during a speech situation.

3

Audience characteristics research study

CHAPTER 3.1

Introduction to audience characteristics

The present study addresses the first two levels of the research model–the audience characteristics, their group dynamic (autonomy), and interaction between audience and virtual listeners. There are two complementary parts to it: the first one is an inquiry into the expert knowledge of public speaking trainers and of experts involved in the design of virtual reality applications for immersive environments; the second part represents an observation of a speech training session held at Technische Universität Ilmenau by a professional trainer.

Seven experts in four fields related to either VR or public speaking were interviewed about three main topics regarding virtual audience adaptation to public speaking and presentation trainings: audience customization, interaction, and group dynamic.

The main premise for this study was that virtual reality applications used in speech anxiety treatment can also be used for non-phobic people, if they match the needs of trainers and trainees. Public speaking experts were required to discuss their expectations with regard to a virtual audience or recall behaviors that turn an attentive audience into a restless one. Experts in VR technology were asked to discuss what is technically doable in terms of audience customization, interaction between speaker and listeners, and with regard to the existing options to achieve group dynamic in a virtual audience setting.

The second part of the study contains a field observation of a one day, five-hour training session with one of the interviewed speech trainers. This put into perspective the role of speech practice within the broad context of skills training seminar and addressed the third level of the virtual audience concept.

A review of virtual reality applications used for speech anxiety treatment in the previous chapter revealed that design can be improved at several levels. Interaction between speaker and audiences can be exam-

ined closer, by looking at the type of feedback speakers receive and at the potential impact of this feedback on speech improvement progress. Moreover, interaction among listeners could also be scrutinized, to see what listeners do during a speech practice session.

In accordance with the research model, audiences have great potential for character diversity compared to what has been reported so far in the anxiety treatment literature, both in terms of demographic characteristics and of behaviors (Verderber et al., 2008; Verderber, Sellnow, & Verderber, 2010). Looking at the roles of trainers, these could be empowered to take control of the simulation and manage not only virtual humans' behaviors, but other characteristics as well, such as looks and outfits.

Both the expert interview part and the observation part of the study are reported here and findings are presented together. The next section will introduce the theoretical background for the main researched topics in the expert interviews. The methodology applied to the expert interviews and observation will be presented next, followed by the results section. The chapter ends with a general discussion on both parts of the study, followed by limitations and suggestions for future work.

CHAPTER 3.2

Theoretical background on real and virtual audiences

Within the audience research model, audience characteristics were assigned a technical valence and a psycho-social one that were both discussed in the expert interviews.

The first researched topic was audience customization. The literature and theory review revealed that, apart from demographic characteristics, cultural, and historical background proposed by communication studies (Verderber et al., 2008, 2010), new audience attributes could be identified from social psychology, specifically from the social impact theory developed by Latane and Zipf (1981). The theory brings into discussion the status of audience members and their relations to the speaker, and the impact these characteristics have on speech performance. Social impact was defined as "any of the great variety of changes in psychological states and subjective feelings, motives and emotions, cognition and beliefs, values and behavior, that occur in an individual, human or animal, as a result of the real, implied, or imagined presence or actions of other individuals" (Latane & Zipf, 1981, p. 343). The level of familiarity of speakers with their audiences was called *personalization* and referred to how familiar are audience members to the speaker, as well as what impact this has on the training, from training experts' perspective. Further audience characteristics, such as demographic characteristics and cultural background were also approached in the interviews.

The second element–interaction–stands out due to its multifaceted nature. On the one hand there is the interaction between the speaker and the virtual audience, which is usually simulated in the case of virtual audiences. Audiences rarely react to the real input from the speaker and are mainly controlled by therapists or programmed in advance to react in certain ways. There are however a few studies that report interactive audience elements, such as Pertaub et al.

(2002), where the audience members could follow speakers with their gaze. Interaction has been so far a characteristic of single embodied conversational agents (e.g., Greta in Pelachaud (2005) or Ada and Grace (Swartout et al., 2010)). The term which represents this process between a speaker and audience is called in this study *speaker-audience interaction*. On the other hand, there is the interaction that naturally occurs among listeners in the same audience, which has been studied in social psychology (Hylton, 1971) and which is less discussed in the virtual reality exposure treatment literature. This type of interaction among audience members was termed here *group dynamic*. This corresponds to the technical attribute of "autonomy" as described in the audience model.

The fourth researched topic–speech practice–emerged at first inductively through the content analysis of the interviews and, as it appeared to yield detailed information into the processes that take place during training session, it was further inquired in the speech practice observational study.

The following chapters cover the theoretical backgrounds for audience personalization, speaker-audience interaction, and group dynamic.

3.2.1   Audience customization

This section reviews a few studies in communication training literature and social psychology in search for relevant audience characteristics that can be added to the virtual audience. An insightful review comes from Buss (1980) who proposes a list of contextual elements, some related to audiences, some to speech location and context that can lead to increased situational communication apprehension. These major elements are *novelty*, *formality*, *subordinate status*, *conspicuousness*, *unfamiliarity*, *dissimilarity*, and *degree of attention from others* (Buss as cited in McCroskey and Daly (1984)). McCroskey and Daly (1984) explain how each factor can impact on situational communication apprehension. Formality and status for instance can impose more rigid rules of what is acceptable behavior in a speech situation, and the higher status party dictates what behaviors are appropriate. Anxiety is high also when the speech situation is novel to someone and they don't have any experience yet on how to deal with it. Furthermore, someone can speak in front of people they know or in front of people

they've never met or who seem to be completely different from them. In this sense, unfamiliarity and dissimilarity can have the same impact on speech anxiety. Standing on a stage and preparing to deliver a speech is a highly conspicuous situation where the speaker receives a lot of attention. However, no attention at all can also increase anxiety level, argue the authors. Beatty (1988) showed that conspicuousness and subordinate status and anxiety levels have a mild correlation. Unfamiliarity and degree of attention were not correlated with anxiety neither in that study nor in a subsequent one (Beatty, Balfantz, & Kuwabara, 1989).

A review of several studies on phobia treatment via exposure therapy in VR showed that the most implemented audience characteristics were audience size, behaviors, as well as speech settings (e.g., a seminar room, a lecture hall) (P. L. Anderson et al., 2005; Grillon et al., 2006; Pertaub et al., 2002; Wallach et al., 2009). Therefore there is great potential research for more factors, such as the ones described above. Many of them can be researched by measuring performance along different degrees of novelty, conspicuousness, or formality and some can also be discussed with trainers who have a overview of training procedure.

One of the topics that has received so far little attention in the design of virtual reality audiences and is tightly related to training procedure is familiarity. People undergo communication skills training to improve the way they present in front of colleagues or managerial levels, improve public speaking before large groups, learn to listen, or present papers to small groups (Di Salvo, 1980). Communication training courses and training seminars take place over days, weeks or even semesters (Cronin & Glenn, 1991), and people get to know each other ultimately. Therefore, it is reasonable to discuss familiarity between speakers and their audiences both during and outside the training sessions, as well as the technical aspects of design supporting virtual humans that look like someone the speaker would recognize.

The present subchapter reviews studies on audience familiarity, status, and audience size with regard to speech performance. Both status and audience size have been already implemented in VR applications, therefore the interview questions focused on familiarity. Nevertheless, status and audience size are briefly reviewed in the following sections and a few suggestions for future studies are made. Furthermore, within

the interview discussions, an extensive amount of information about general audience characteristics emerged, and audience size and status are also reported in the results chapter.

3.2.1.1   Familiarity

In a person's history of speeches, they may have had to face audiences with people they've maybe seen or even known before as well as audiences of complete strangers. Audiences can consist of friends, acquaintances (Maclntyre & Thivierge, 1995) or of known or unknown supervisors (J. L. Cohen & Davis, 1973).   The level of familiarity between speaker and pubic can result in different anxiety levels and variable willingness to speak: people predicted various anxiety levels based on whether they imagined speaking to friends, acquaintances, and strangers, depending on the audience friendliness or unfriendliness levels, and on speaking context (academic, social, and professional) (Maclntyre & Thivierge, 1995).   Results showed that they preferred speaking to familiar audiences, friendly audiences, and audiences in professional contexts (Maclntyre & Thivierge, 1995). In both academic and social contexts, they were more willing to speak to pleasant strangers than to unpleasant friends, and in professional contexts they preferred to talk more to pleasant friends than to unpleasant strangers (Maclntyre & Thivierge, 1995).

Familiarity between speaker and audience can correlate with performance levels, as Butler and Baumeister (1998) report. These authors showed that there was a discrepancy between performers' perceived stress towards a familiar audiences and how well they actually performed in front of that audiences. People performed tasks slower and were more cautions in front of friends than in front of strangers, however they perceived marginally less stress from the friendly audience and also less distractions from known audiences (Butler & Baumeister, 1998). Finding a relation between performance and familiarity is an encouraging factor for devising a virtual audience meant to help improve speech performance. Future studies could measure whether training in front of known people is more effective than in front of total strangers, and also what is an optimal mix between known and unknown people within the same audience.

The other way to look at familiarity is through training procedure: people who attend together public speaking seminars get to know each other and it is yet unknown how this affects training effectiveness in

time. Therefore, familiarity with audiences, both outside training and within the training context emerged as a relevant topic of research and was further inquired in the interviews.

Several speech contexts were mentioned above (academic, social, professional) in combination with known and unknown audience members (Maclntyre & Thivierge, 1995), and this hints towards the status of potential listeners. In one study, audiences were used to address public speaking performance, and they consisted of novices and experts in public speaking (Hilmert, Christenfeld, & Kulik, 2002). Subjects manifested higher physiological responses (e.g., increased heart rate) to expert audiences than to novices, but these effects applied to subjects who saw themselves as being less confident with public speaking (Hilmert et al., 2002). An important finding with regard to audience expertise was that novice audiences had little impact on cardiovascular responses of speakers, independently on how confident that speaker considered himself/herself to be. Researchers explained that this could happen because speakers do not consider the feedback of novices important in comparison to that of experts.

Audience status was previously implemented in anxiety-eliciting situations, such as a job interview with a virtual character (Grillon et al., 2006). Since familiarity has yet to be implemented at all in the reviewed literature on virtual audiences, trainers were required to discuss their opinions and share their experience on the role of familiarity with an audience in relation to trainees and how this element can be used in the training procedure. Virtual reality experts were required to discuss the technical options for creating virtual audiences that look like and behave like someone a speaker would know. Through the open-ended structure of the questions it was expected that data about audience status would also emerge in the discussions.

3.2.1.2   Audience size

Another audience characteristic discussed in relation to performance is audience size. People with high anxiety levels act differently than people with low anxiety levels and this happens in the company of different audience sizes (McKinney, Gatchel, & Paulus, 1983). Less anxious people can actually report feeling more comfortable with more audience members compared to more anxious speakers who are negatively aroused by numerous audiences. McKinney et al. (1983) reported such findings for three conditions: no audience, an audience

of two, and one of six people, where the latter caused least negative feelings to non-anxious people. Jackson and Latane (1981) measured nervousness of single performers in front of a large audience of 2.500 members. People reported feeling seven times more nervous performing in front of this large audience compared to a usual 20-people university class. However, anxiety grows as a function of audience size, but slower (Latane & Harkins, 1976). A threshold where anxiety levels became stagnant was not reported. The study proved also that audience status impacted anxiety levels, with higher status audience members exerting more pressure on the performer. The drawbacks of the study are that the audience was static and the performance task was imaginary: subjects had to imagine memorizing a poem and reciting it in front of various audiences showed as photographs.

Virtual audiences in anxiety treatment have been designed to accommodate various audience sizes, with usually small audiences. Pertaub et al. (2002) used positive, negative, and neutral audiences made of eight formally-dressed male virtual humans; North et al. (1998) and P. L. Anderson et al. (2013) used audiences that could accommodate up to 100 people.

Audience size is an element that has been customized for virtual audiences and the interviews looked more at audience diversity–an important aspect for large virtual gatherings that hasn't been explicitly addressed so far in virtual training procedures. It was discussed indirectly though, by asking experts how it is technically possible to create large audience in a believable manner.

3.2.2 Interaction

Interaction in a speaker-virtual audience communication setting is a topic at the intersection of various research fields. From the speech format perspective supported by the communication studies, public speaking is a discourse taking place in front of a group of people, where the information is usually prepared in advance and presented in a limited time slot (Lucas, 2012). Here, the verbal speaker-audience interaction is regulated and usually takes place at the end of the speech, in the question and answer sessions (Lucas, 2012). Nevertheless, there is a continuous exchange of information between these two actors and this

contains more than just verbal communication, as the social psychology perspective points out next.

Public speaking can be considered also from the social psychology perspective, as setting for *face-to-face interaction*, with speakers addressing a group of people gathered to listen to a live speech. Face-to-face interaction can be defined broadly as "the conscious or out-of-awareness exchange of behavioral and nonbehavioral, sensible and intelligible signs from the whole arsenal of somatic and extrasomatic systems (independently of whether they are activities or nonactivities) and the rest of the surrounding cultural systems" (Poyatos, 1985, p. 111).

According to Poyatos (1985), face-to-face interaction entails two components: (a) the internal component, which contains any active or passive elements that can be exchanged between interaction partners, such as for instance language and (2) the external one, which represents elements that impact on the interaction, but stem from the environment where interaction takes place (e.g., environmental sounds, physical space where the interaction takes place). He continues by assigning interaction between speaker and audience to a basic triple structure that contains language (content of information), paralanguage (voice modifications such as pitch or volume), and kinesics (facial and body movements and postures). Speakers and audiences use language, paralanguage, and kinesics to interact with each other, and the way this process takes place is regulated first by the speech discourse format that allows speakers to give verbal input and listeners to respond verbally usually at appointed times (e.g., in a question and answer session at the end of the speech or presentation) (Poyatos, 1985). How the interaction actually takes place and what it contains was subject of inquiry in the expert interviews. The speaker-listener interaction is discussed in detail in the following section.

The third perspective on speaker-virtual audience interaction is given by the human-computer paradigm. Speakers would have to address the virtual humans in the audience, and this brings into attention the interactive capabilities of virtual characters. The first prerequisite for interaction is autonomy, as discussed previously in the section on virtual humans characteristics. A virtual human can respond in a natural manner to a human if it perceives its environment, if it can take decisions to act, and if it can enact cognitions and emotions both verbally and nonverbally (Kasap & Magnenat-Thalmann, 2007).

3.2.2.1   Speaker - listener interaction

Human behavior is constantly communicating about a person, independently of how intentional the actions are. As Watzlawick, Beavin, and Jackson (1967) state, "one cannot not communicate" (p. 277). They define a single communication unit as "message" whereas a series of exchanged messages between people is called "interaction".

During a speaker-listener interaction, the listener performs specific behaviors, called backchannels (Yngve as cited in Duncan Jr. (1972)). Backchannel is a concept that stems from social psychology research on face-to-face conversations and signals that the listener is engaged in the interaction with the speaker (Duncan, 1974). In dyadic conversations, several verbal backchannels signals occur, such as "m-hms" or "yeah, yeah", sentence completions where listeners complete a sentence started by speakers, requests for clarification, and brief restatements of what the speaker previously uttered (Duncan, 1974). Among the identified nonverbal backchannels are head nods and shakes (Duncan, 1974) as well as smiles (Lawrence, 1979). The role of these backchannel is to signal engaged listenship (Lambertz, 2011), encourage the speaker to continue the conversation, and also suggests that the listener does not wish to take the turn in the conversation (Duncan, 1974). For presentations and public speaking audiences, the same engaged listenship manifests through displayed attention (Wilson & Korn, 2007) or immediacy markers (Baringer & McCroskey, 2000). In artistic or political contexts, it is expected from audiences to engage in public displays of their reactions, such as applauding or cheering (Barkhuus & Jørgensen, 2008; Heritage & Greatbatch, 1986).

One of the main goals of training presentation and public speaking skills is to make speakers aware of the interaction between themselves, message, and audience (Turk, 2004), as well as to manage backchannels and improve a speaker's performance in time (Beebe, Mottet, & Roach, 2012). Moreover, in advanced public speaking classes, apart from speech practice, students are required to learn how to give and receive feedback (Levasseur et al., 2004). This feedback is a conscious and constructive response to how the speaker performs and can come from both trainers and fellow trainees at the end of the speech (Levasseur et al., 2004). By this it differs from backchannel responses which may occur during the speech, without interrupting it.

3.2.2.2   Interactive virtual humans

Human interaction is a starting point in devising virtual audiences. The
next section covers both the technical attribute of autonomy (Kasap &
Magnenat-Thalmann, 2007) from the first level of the research model
as well as interaction as information exchange among speakers and
listeners, at the second level of the model.

Interaction with virtual humans has long been a topic of interest in the
design of virtual reality applications. As Elkins, Derrick, Burgoon, and
Nunamaker Jr. (2012) pointed out, the "natural progression of human
interaction with machines is leading to systems that can automatically
and unobtrusively assess human states, such as anxiety, satisfaction,
boredom, and credibility, and interact naturally with human users
based on these assessments" (p. 579). Virtual characters have become
increasingly common and serve in a wide range of applications, from
museum guides (Ada and Grace by Swartout et al. (2010)), and local
guides Huang (2009), to animated pedagogical agents in virtual learning
environments (Johnson, Rickel, & Lester, 2000). As their application
range diversifies, virtual humans need to become more and more
sophisticated in their interaction with humans.

Of particular importance for realistic interaction is emotion simulation
and display. Emotion simulation has received extensive attention in
the design of interactive agents that can not only read emotions, but
also display facial expressions and body movements that denote their
affective state. Such models rest on emotion appraisal theories (Rank
& Petta, 2012; Scherer & Ellgring, 2007). The componential appraisal
helps a character capture and evaluate the whole environmental
complexity around it and adapt to the changes in this environment
(Scherer & Ellgring, 2007). This model is particularly relevant for
interactive agents. It allows characters to be flexible and adapt
outside their virtual environment to the interaction with a user and
the unpredictability of real life situations. Hence, characters not only
process information and respond with proper gesture and posture,
but their response denotes also emotional states, just as it would
in a human. Embodied Conversational Agents and Special Purpose
Embodied Conversational Intelligence with Environmental Sensors
which are usually used in one-to-one interaction with people make use of
various sensors that support autonomous action up to a certain degree
(Elkins et al., 2012; Pelachaud, 2009).

Real time feedback requires sensors for body movement detection such as cameras, magnetic motion capture, data gloves, head trackers, and microphones for voice detection (Huang, 2009).   These are backed up by a system architecture that usually varies with the application purpose and integrates blocks specific to each virtual character.   For instance, the architecture of one of the first embodied conversational agents–REA–the virtual real estate agent contains an understanding module, a decision module with interactional processing, propositional processing and response planner, and a generating module which gives the character's reactions (Cassell, Sullivan, M., & Churchill, 2000). A different agent has a different architecture, such as the relational agent that provides social support for healthy but isolated elderly. The architecture, called DiscoRT contains perceptors (face, motion, or speech recognition), schemas, behavior realizers and resources (e.g., hand, face, voice) (Nooraei, Rich, & Sidner, 2014). However, autonomy and genuine interaction are rarely the rule of thumb in phobia treatment studies, where interaction between speakers and audiences has been usually simulated through the Wizard of Oz technique, by having someone (a confederate) control how virtual humans react to a speaker (Grillon et al., 2006).

As already seen, interaction can be discussed from various perspectives that cover broad fields of research.   In order to grasp its full meaning in the context of training, experts were encouraged to give their own definition of interaction in the context of a public speaking virtual audience.   Through this method they could not only provide their expertise on what interaction in this context is, but also establish what is important from their point of view with regard to interaction and how this can be technically achieved.

3.2.3   Group dynamic

Group dynamic refers here to those actions that group members undertake when they interact with each other within the group and with the performer.   This element corresponds to "autonomy" in the first level of the research model (Kasap & Magnenat-Thalmann, 2007). Kurt Lewin (1951) called group dynamics "the actions, processes, and changes that occur in social groups" (Forsyth, 2005, p. 16). Within the context of a virtual audience proposed here, group dynamics refers

to the actions that occur among listeners, as well as to the way such actions could spread dynamically from one member to another.

Human gatherings have been studied in terms of their dynamic, particularly with regard to how they move and organize in space or how a small audience invites more people to join in until a saturation point is reached, both in real life situations and in virtual simulations (Forsyth, 2005; Knowles & Bassett, 1976; Sharma, Otunba, & Han, 2011). In comparison to spontaneous groups that form on the street (e.g., queues, mobs, or riots), organized gatherings that occur at sport events, concerts or public speaking events (e.g., conferences, seminars, lectures) form audiences or conventional crowds (Forsyth, 2005). Forsyth (2005) defines them as weak associations, because they represent groups that form spontaneously, last little, and have permeable boundaries. Whereas conference audiences could fit into this categorization, other groups that can serve as virtual audiences for presentation and public speaking skill training include more structured groups such as seminars or work groups of people who already know each other.

Spectators or listeners are subject to collective behaviors, which represent "the actions of a group of people who are responding in a similar way to an event or situation" (Forsyth, 2009, p. 508). Within these audiences, people perform several behaviors in accordance with the norms of what is or isn't acceptable within the group and in accordance to the setting of the observed event (e.g., clapping, cheering) (Forsyth, 2005). In the end they disperse in orderly fashion after the event is over (Hollingworth, as cited in Forsyth (2009)).

In performing these setting-determined behaviors, audience members look both to the performer and to each other. This type of behavior is called contagion and it represents the tendency of people to follow the behaviors of other people in the same group (Wheeler, as cited in Forsyth (2009)). One example of contagious behaviors is the occurrence of applause and laughter (Kowal, 2009). Applauding can occur both spontaneously, as reaction to a performer when several audience members respond simultaneously, or it can occur as response to mutual monitoring, when only a few start it and the rest follow (Clayman, 1993).

Other types of contagious behaviors occur as "chameleon effect": when people interact with each other, they tend to sometimes match the

reactions of the persons they are with, such as posture changes or gestures and such postural mimicry tends to occur when people are other-oriented (i.e., try to be liked or seek to connect with others) (Chartrand & Bargh, 1999; Knapp & Hall, 2010).

Group members can be subject to coordinated audience reactions, that are either based on independent decision-making or on mutual monitoring (Clayman, 1993). Both sources of reaction are relevant for the design of virtual audiences. The independent decision-making process through which a listener uses backchannels to react can be assigned to the interaction between a speaker and audience members and has been discussed above in the chapter on interaction. The mutual monitoring processes that determine behaviors such as behavior matching or behavior spreading can be further researched and implemented in the design of a virtual audience, to fully grasp the audience behaviors that occur in real life when people sit and listen to a speaker.

3.2.3.1 Virtual groups and their group dynamic

The study of virtual groups in VR has been mainly focusing on crowds and simulation of crowd movements in large spaces. A few examples contain the simulation of a virtual crowd, such as the one in the ancient setting of the Colosseum which was meant to test the various hypotheses about the architectural qualities of the monument (Gutierrez, Frischer, Cerezo, Gomez, & Seron, 2007). With the help of a crowd of eight thousand virtual agents, researchers were able to identify for instance several bottlenecks in the distribution of people in the building. In another study, virtual humans were used to simulate emergency evacuations on board of planes to train flight attendants for emergency situations (Sharma et al., 2011).

Crowd simulations refer mainly to the coordination of movements of people in open or enclosed spaces and makes use of goal seeking or wandering algorithms (Sharma et al., 2011) and algorithms of collision avoidance and path planning (Ulicny & Thalmann, 2002). Such complex behaviors require that virtual agents are autonomous and perceive the environment, are aware of their goals within that environment, are able to take decisions based on the knowledge as well as occupy and change the environment they populate (Gutierrez et al., 2007). A model that solves this problem is the spatial interaction model discussed by Benford and Fahlén (1993) and applied to multiuser virtual reality systems. The researchers suggest that space is used

to mediate interaction and objects interact in the virtual space through *aura, awareness, focus, nimbus*, and *adapters.* The aura marks the presence of an object within a space and it always accompanies an object. A virtual human, for instance, carries an aura. Another virtual human, through their awareness function, can perceive this aura and therefore the presence of the previous one. Awareness levels are not inherently symmetrical, and each actor can chose whether to become aware of the other or not. This is further mediated by focus and nimbus. These are elements used to negotiate the mutual and potentially non-symmetrical awareness between actors: the awareness level of an object A on object B is a function of A's focus on B and B's nimbus on A (Benford & Fahlén, 1993, p. 112). Once interaction is achieved through these elements, it can be further shaped with the help of adapters which can modify the aura, focus, and nimbus of an object. In concrete terms, a virtual character might use a virtual microphone (adapter) to make himself/herself heard by other characters who were not previously aware of him/her (had no focus). Such a model would respond to the needs of a virtual audience because it offers conceptual solutions on how to make each character of a group aware of the others and how to influence this level of awareness through various visual and auditory stimuli.

In contrast to the study of spontaneous groups movements that occur in public spaces or during emergency evacuations, an audience has far less action options, which are mainly dictated by the location and the seating options: they come in, take a seat, and leave. The way seating constellations look like, who do they seat next to, how do people seated to each other interact with each other can be observed and understood only if listeners are asked about their personal choices. In order to simulate behaviors that rest on complex personal backgrounds and not only on rules of avoidance in space, one has to look at psychological models in the behaviors of virtual agents which can include a virtual agent's motivations, emotions, or personality (Silverman et al., 2005), and even common history with fellow audience members.

Another aspect of group simulation and dynamic is behavior diversity. If people are dynamic and can influence each other, how should then virtual humans' animations be constructed so that they don't appear copied from each other all over again? Variety is considered key in enhancing the realism of multiple agents simulations and it stands for the capacity of each simulated virtual human to differ from another

(Ulicny & Thalmann, 2002).   The two researchers propose several levels of variety to deal with implementing differences of behavior into the behavior models of virtual crowd members. At *level zero* there is no variation at all and the system is using only one solution to fulfill a given task. In terms of behaviors this would mean that all virtual agents would behave the same way. It can be argued that this approach hinders audience realism. With *level one* there is a finite number of solutions, which would translate into implementing various behaviors for the virtual humans (various behaviors that occur repeatedly), whereas *level two* would mean that there is an infinite variety of solutions and the system has to find a way to assign them in a randomized manner. Ideally, a VA would find itself either at level two or three.

Two relevant topics emerged from the literature research that are further inquired throughout the expert interviews: (1) what types of group behaviors occur in the real audiences training experts have worked with and (2) how can behaviors be assigned to several virtual humans without making them too repetitive throughout the whole listener group.

3.2.4   Summary

The scope of this study was to identify relevant aspects that can serve the design purpose of a virtual audience. Three topics were identified in literature that can help improve virtual audiences and enhance their applicability to training scenarios. These topics can be found in the audience research model at the first level, which includes audience characteristics, and at the second level, which includes interaction between speaker and audience. Table 3.1 summarizes the three topics and their components addressed in the study.

Table 3.1
Frame of research

Topic of research	Components
Audience customization (Hilmert et al., 2002; Jackson & Latane, 1981; Maclntyre & Thivierge, 1995)	Audience familiarity;
Interaction (Levasseur et al., 2004; Lucas, 2012; Poyatos, 1985; Watzlawick et al., 1967)	Speaker-listener interaction; Interaction as feedback session; Human-computer interaction.
Group dynamic (Forsyth, 2009; Ulicny & Thalmann, 2001, 2002)	Contagion; Variety.

All in all, these three topics discussed here together with their components cover a fraction of all potential design elements that can be implemented into a virtual audience. The decision to focus only on these themes is justified by the fact that the audience is conceived for public speaking and presentation skills training and its capabilities should be first relevant for this endeavor. Another design element that can be studied in the future is flexibility of the application, such as making it available for several portable devices. Sound would also be a relevant topic of future research in order to understand and assess its impact on speech performance. The cognitive capabilities and content-related feedback could also be relevant in the speaker-virtual audience interaction, particularly for advanced training where speakers need in-depth feedback on the content of their speech.

CHAPTER 3.3

Methodology

The goal of the two-parts study presented here was to reveal audience characteristics and what roles these play during public speaking training, to show what technological opportunities and barriers exist in implementing these characteristics in a virtual environment, as well as to reveal the training procedures in a speech seminar. Following, the research design is briefly introduced. Both parts of the study have an exploratory approach and belong to descriptive research (de Vaus, 2001). They investigate without having previous explicit expectations (Schutt, 2014). In order to acquire the data for the first part of the study, a semi-structured, in-depth expert interview was conducted with several experts from various relevant fields of research related to virtual reality and public speaking training: design and system architecture, programming, psychology, and public speaking and presentation training.

There are several methods used to elicit expert knowledge, such as analysis of familiar or special tasks as well as interviews (R. R. Hoffman, 1989). Task analyses help chart reasoning processes of subjects and acquire insight into decision making processes and are considered most suitable for revealing what experts know and how they apply the knowledge. Interviews on the other hand can give account of experts' beliefs (R. R. Hoffman, 1987) and for this reason they were the data collection method of choice here. The semi-structured expert interview has several advantages: (1) it offers relevant information in a brief amount of time, (2) it gathers expert knowledge in complex fields that require in-depth information search, (3) it is a method that can be applied in the early development phases of a project, by tapping into expectations of potential users (e.g., public speaking and presentation trainers).

The expert interviews were completed by a field observation of a speech practice session with one of the public speaking trainers. Participant

observation is an anthropological method, which allows researchers to tap into processes that are difficult to describe otherwise and usually serves as starting point for further investigations, such as interviews, questionnaires, and theory and hypothesis development (DeWalt & DeWalt, as cited in Kawulic (2005)).   The decision to conduct the observation came after the expert interviews and it served as complementary method for gathering information about the practical part of a training seminar (the third level of the research model).

The observation comprised one training session of five hours. Based on the results, it is possible to understand how speaker-audience interaction occurs and also to envision the role of the virtual audience within the broader setting of the public speaking and presentation training.

The observation was organized based on the principles of ethnographic studies, with the observer as participant (Gold, as cited in Kawulic (2005)). The presence of the observer has been acknowledged by the training seminar participants and each participant as well as the trainer signed a consent form. With regard to observation type, this was a focused observation (Werner and Schoepfle, as cited in Kawulic (2005)), as the researcher already knew that the observed meeting constitutes the major speech practice session of the whole seminar. The observer took field notes of three aspects such as: the unfolding of the seminar with the discussed and reviewed topics, the actual speech practice, and the types of feedback and interaction between trainees as speakers and trainees as audience.

3.3.1   Research questions

Three main topics guided the research initially: audience customization, interaction, and group dynamic. A research question was assigned to each of them. The fourth question emerged after the expert interviews, when it became clear that speech practice sessions are a topic that needs separate attention, because it can reveal important information about audience roles within the training concept.

Speaker's familiarity with an audience. The research question RQ1 was: *What is the role of audience customization in the training and how can this be technically acquired?* To answer this question public speaking and presentation trainers were asked primarily about the role a known

and unknown audience plays in the training program and whether they implement this audience typology in the training sessions, as well as about audience numbers and status. Technology experts were required to discuss these elements from the technical perspective and consider them in the context of 3D immersive virtual environments.

Speaker-virtual audience interaction. The research question RQ2 was: *What is the importance of interaction between speaker and audience in the training process and how can interaction be acquired?* The first half of the question targets the opinions of the trainers and the emphasis they put on the interaction between the speaker and the audience: how is interaction described by experts and how are trainees taught to achieve it? The second half addresses the experts from the technical filed, who can assess how one can achieve such interaction. These experts were required to describe methods to endow virtual humans with interactive capabilities and also to assess the effort needed to achieve this.

Group dynamic. The research question RQ3 was: *What group dynamic behaviors occur within a real audience and how can they be implemented in a virtual audience?* Experts in both training and technical fields were required to discuss group dynamic, to assess its importance in the training process and to share their knowledge on how to implement dynamic and diverse behaviors from the technical perspective.

Speech practice. The research question RQ4 was: *How does a training session look like and how does it accommodate the speech practice and speech feedback?* The question targeted the actual speech practice setting, which is little discussed in the communication training literature. There is common agreement among researchers that there are practice sessions where trainees get to apply what they have learned in theory (e.g., Beebe (2007)). Yet the hard facts about how such sessions look like are scarce. Lastly, it was expected that the more is known about the speech session, the more precise the virtual audience can be devised and applied in the broad communication training process.

3.3.2   Expert interview guideline

To adapt to the exploratory nature of the research, a semi-structured interview guideline was developed.   The semi-structured interview has the advantage of offering a structure to the discussed topics without restricting the information flow (Bortz & Döring, 2006). The interview guideline used with the experts contained three main areas of discussion, for each of the topics of interest. Each category contained several questions tailored to the needs of the expertise of the interviewee.   For instance, public speaking trainers were asked to provide their own definition of interaction between a speaker and audience, as well as its importance in training altogether.   On the other hand, technology experts were asked how it's possible from the perspective of existing technologies to foster interaction between a speaker and a virtual audience.   There are four separate interview guidelines that addressed designers, programming experts, public speaking trainers, and a VR specialist who has worked in the development of a virtual audience for fear of public speaking treatment.

Apart from the main discussed topics in the study, the designers' view was meant to present also elements of system architecture, such as how to plan and design a system that supports a virtual audience, in order to propose a model that can be implemented at TU Ilmenau.   For the present analysis though, only the information relevant to audience personalization with familiar faces, interaction, and group dynamic was reported. The interview with the specialist in virtual audiences used in VRET procedures proved also very useful from a technical perspective, as the expert himself had a computer science background.   Therefore he and several other virtual reality experts were merged together into a common group called VR experts. Their interviews were analyzed in a complementary manner, in order to provide a coherent view on the technical possibilities of a virtual audience for communication training purposes.

This type of qualitative interview allows for complex data collection with regard to the many facets of personal experiences of the interviewees. In line with Wiedemann (as cited in Bortz and Döring (2006)), who proposes several types of subjective experiences, the experts were asked about personal experiences with virtual reality applications, hard knowledge about how virtual reality works, what's technically possible to design, but also how they would imagine a virtual audience.   Last,

public speaking trainers who have never interacted with a virtual reality system before were given the opportunity to imagine using a virtual audience for the training and to describe the features it would have. They were asked to think of ways of using the application and the virtual audience in any way they would possibly want (both in terms of simulated content as well as in terms of user interface). They were encouraged to speak about anything they would consider relevant for the training and to describe in detail all their expectations from the virtual reality training scenario and user interface. Although this is a criticized practice in design (Jerald, 2015), experts were asked about these facts to explore their general expectations on a product that is still in the prototyping phase.

The open ended questions and the semi-structured nature of the interview left a lot of room for spontaneous questions. The questions were pretested with two persons familiar with the field of virtual reality from Technische Universität Ilmenau.

3.3.3 Sample

For the expert interviews, the sample contained seven experts. The creation of the virtual audience and of the setting is a multidisciplinary endeavor and it needs to draw on the knowledge of people involved in these fields. Ultimately, the training scenarios need to combine answers to the research questions presented above and merge what's important from the training perspective with what is actually doable from design and programming perspectives.

The selection of experts was made based on following criteria: (1) academic and/or publishing record in the virtual reality field, (2) affiliation with organizations that are active in the field of virtual reality, (3) affiliation and professional activity in public speaking skills training. Based on these criteria, a list of potential experts was created and in total seven agreed to participate: three experts in design of virtual reality applications and one participant in programming of virtual reality applications. One expert has previously worked with therapists in the development of a virtual audience for fear of public speaking. These five represent the VR technical angle to the virtual audience design.   The remaining two experts are public speaking and presentation trainers.   These had no practical experience with

Table 3.2
Experts' backgrounds

Assigned domain	Expert	Expertise and name initials
Design of virtual reality applications	1.	Computer science, human-computer interaction and interactive virtual humans (RK)
	2.	Architecture and computer science, VR spaces design (MNA)
	3.	Computer science, graphic animation and group dynamic (DT)
Programming and visualization	4.	Computer science, 3D visualizations (TK)
Psychology	5.	Computer graphics, virtual reality, public speaking (BH)
Public speaking	6.	English language, public speaker trainer for competitions (RM)
	7.	Linguistics, pubic speaking trainer (CS)

virtual reality applications of any kind. All participants, whose field of expertise and name initials are presented in Table 3.2, are active in their field and have proven academic track record. Out of the seven experts, two were women.

The composition of the experts sample is relevant because it contains trainers, who are part of the main stakeholder group benefiting from the virtual reality training scenarios. Their expert knowledge is particularly important because they can discuss what's relevant from their training perspective. The other experts can also share their experience with virtual reality applications and point out opportunities and risks in the design of virtual audiences.

For the field observation, which took place on day three (the practice day) of a training seminar organized at Technische Universität Ilmenau, the sample contained ten participants and one trainer. The tenth participant arrived in the second half of the seminar, when he was briefed and given the consent form. There were eight men and one woman. Except for the trainer who was German, all participants were international students. Age range lies approximately between early 20s-mid 40s years old. Age was discussed with a few people but was not inquired on a separate questionnaire.

The data from interviews and the observation is presented under anonymity.

3.3.4 Data collection and content analysis

Interviews were taken over a period of two years. Technology experts and all but one public speaking expert were interviewed in June-September 2012. The last expert on public speaking was interviewed in December 2014. All but one of the interviews were conducted via Skype. One interview was conducted face to face. All interviews were recorded, transcribed, and uploaded into MaxQDA for content analysis. All but two expert interviews were conducted by students in a Bachelor research seminar. The other two were conducted by the researcher.

For the observation, field notes were taken by hand and a memo file was later compiled which contains all the relevant facts that were noted down.

The interview guidelines were split into the three major researched topics: customization, interaction, and group dynamic. The open-ended nature of the questions allowed for the transcribed information to be analyzed beyond these three topics which were also the initial major analysis of categories.

The content analysis used the coding method (Gläser & Laudel, 2013), both as indexing topics and as conceptual device, through which connections between categories and their underlying concepts could also be analyzed (Schreier, 2012). As Schreier (2012) further points out, coding is appropriate where several groups are compared, such as the current study where experts from various fields discussed common

issues. The choice for coding over qualitative content analysis was grounded in the following facts:

- the research questions were exploratory;

- the interviewed experts came from different work fields and work practices which leads to

- expected differences in the meanings assigned to the discussed concepts.

Due to these factors, it was hard to establish from the beginning clear directions for data interpretation and to use a predefined codebook to manage and analyze the data, as required by the qualitative content analysis (QCA) practice. Moreover, the information has a very high interpretative value, and data reduction procedures used in QCA could lead to the omission of relevant information.

In the first step of the analysis, the data was split into the three main researched topics. They also received definitions and coding rules in the MaxQDA coding platform where the analysis was conducted. These represented the major categories from where the analysis began.

After pooling all data into the content analysis software and assigning it to the categories in the codebook, the texts were further analyzed inductively. In the first cycle coding, the initial coding method was applied (Saldaña, 2010) to help structure the information and identify potential relevant themes for the purpose of the study. This step was followed by a second cycle coding process, through which all data was reorganized and reanalyzed (Saldaña, 2010). Through this method it was possible to identify properties of categories, causes for discussed events and conditions in which they occur.

The analysis process continued in an iterative manner and new categories and subcategories emerged. Several category revisions occurred until all relevant information got assigned throughout the whole category system.

3.3.5 Ethical considerations

The observational study was conducted in agreement with the American Psychological Association principles of ethics (D. Smith, 2003). Credit is given as follows: the study was conceived and partially conducted together with a Bachelor seminar group at the Institute of Media and Communication Sciences, at TU Ilmenau, in the summer semester of 2012. The guideline for the interview was conceived together with the students and pretested by them. Seminar participants found and contacted all VR specialists and carried the interviews with them. The other two interviews with the public speaking experts, as well as the speech session observation were carried by the researcher alone.

Interviewed participants were contacted in advance per email and were asked to voluntarily participate to the study. Upon agreement, they were briefed about the purpose of the study and were required to give their verbal consent to participate. They were ensured that the data will be handled with great confidentiality and only for academic purposes. The contact data of the researcher was also provided, in case they had further questions. Participants were not given any financial incentive.

For the speech practice observation, the researcher obtained the verbal consent of the trainer to present the study to the participants and to ask for permission to join the seminar and observe the session. Participants were briefed about the purpose of the observation and were given a written consent form. No personal identification data was recorded, and the observation was conducted with paper and pencil. No filming or photographing occurred. The contact data of the researcher was also provided, in case participants had further questions. Participation was voluntary and no financial incentives were given.

CHAPTER 3.4

Results

The results reported here contain findings from both the expert interviews and the speech practice session observation. Timewise, the observation took place after the interviews and was conducted as a consequence of preliminary interview findings, specifically on real life training practices. This led to the adding of new categories to the content analysis.

Initially, three topics served as the main categories of analysis (familiarity, interaction, group dynamic) to which a forth one was added after the observation–real life training practice). However, the body of data yielded information beyond the four research questions formulated initially. When this is the case, the grounded theory approach suggests reformulating the research questions to cover the newly discovered data (Charmaz, 1996). The data which could not be subsumed under either of the research questions formed three other separate themes. These ones represent complementary information about virtual audience characteristics, the setup of a virtual audience, and the limitations and risks of virtual audience design. All this information adds to the audience design concept and to the scenarios proposed in the last study of the dissertation. The seven topics reported in separate sections are as follows:

1. Audience customization. This category contained the findings that addressed the first research question on how to make audiences look familiar to trainees. The theoretical chapter contained, apart from personalization, also audience size and status. These two latter elements are reported under general audience characteristics (see point 5).

2. Interaction. The second topic represents the findings on how interaction looks like in the feedback session, from a one-to-many perspective, and between a human and a virtual counterpart.

3. Group dynamic. The third topic gathered all specific data on how group members behave and interact with each other in real life, as well as on how to obtain group behaviors and variety in virtual reality. Variety is reported under the larger topic of virtual audience characteristics (see point 5).

4. Real life training. The category contains detailed information about (a) who participates in training sessions, (b) how does the training procedure looks like, and (c) training session formats (e.g., in a public speaking club).

5. Virtual audience characteristics. This topic contains several subcategories that cover information about what trainers expect and need from a virtual audience to what the audience behaviors and demographic characteristics should be.

6. Virtual audience setup. This category addresses trainers' expectations with regard to the user graphic interface as well as other references to GUIs used in phobia treatment.

7. Limitations and risks of virtual audience design. This topic sums up some of the design limitations and risks designers came across in similar projects, which should be avoided in the future.

The quotations used here are simplified and corrected both grammatically and syntactically and some terms are used interchangeably, such as "trainee" and "patient" and "trainer" and "psychologist". Certain experts used the term "patient" when referring to the subjects in training or treatment session and the term "psychologist" for the role assigned here to the trainer or coach. The persons are quoted in original, with all the original denominations they used. All experts are marked in the text with their name initials, followed by two acronyms: TE for public speaking training experts and VRTechE for VR technology experts.

The findings are presented selectively in tables, with relevant codes and interview excerpts. The tables contain three columns. The first column represents the used codes throughout the analysis. The second column comprises interview excerpts with quotations by the experts to exemplify the codes. The third column contains suggestions for putting the findings into practice for virtual audience design.

3.4.1   Audience customization - Familiarity

The first topic proposed for the virtual audience concept represents a particular quality of audience members, that of looking familiar to the speaker. The research question that guided the discussion was: *What is the role of audience customization in the training and how can this be technically acquired?*

Substantial literature discusses the demographic qualities of audiences and their impact on a speaker, as well as the impact of familiar faces on anxiety levels. A less researched factor that is highly recurrent especially in long training programs is the fact that, at some point during the training, people will start getting to know each other. This means that, sooner or later, people will present in front of known people and this makes it difficult to simulate real-life speech situations with diverse audiences. One training expert pointed that out clearly: "these people are colleagues in the same public speaking club and they kind of tend to make it funny" (RM_TE). Table 3.3 illustrates the points discussed with the experts.

Public speaking and presentation experts acknowledged the fact that each speaker has personal preferences when it comes to addressing people they know or don't know. This could make it hard for a trainer to establish a rule of thumb with regard to the training audience they would prefer to use. For some speakers "it's better to have a familiar situation or familiar persons or friends" (CS_TE). The expert explains that the difference between having known and unknown audiences lies in being aware of the influence that strangers can have upon the speaker: "it's important in the way of being aware that the situations are different". Due to the seminar composition, where people get to know each other, each speaker has the option to image speaking to a different audience than the one in front of them. On the one hand, this can be advantageous since every speaker can decide freely who their listeners could be. On the other, people might have different imagination resources and might not be able to get into the role and convince themselves completely that they speak to someone else than their colleagues from the speech training seminar. With a virtual audience at hand, it would be possible to translate the qualities of an imagined audience (e.g., their looks, outfits, roles) to a virtual one which could be customized each time.

Table 3.3
Familiarity with the audience

Code	Interview excerpts	Conceptualization for virtual audience design
personal preference	"It depends a little bit on the personality of the speaker." (CS_TE); "Some speakers say 'no, for me it's OK if they (listeners) are really foreign."' (CS_TE).	customize virtual humans' looks to accommodate familiar and unfamiliar faces
imagination	"the imagination helps" (CS TA); "[...] OK, just imagine we all are here your family." (CS_TE); "We all here are strangers and we judge you, because we have to, because you make a dissertation or whatever" (CS_TE).	serve as alternative to imagination exercises
performance	"The more they have outings outside of their safe environment, where they can practice and get feedback, the better, the more accustomed they get to speaking in front of complete strangers" (RM_TE).	various demographic characteristics of virtual listeners and outfits; expressive capabilities through verbal and nonverbal behaviors.

The other public speaking expert approaches the matter differently and mentions the advantages of speaking in front of unknown audiences for the overall performance. Since she is also training pupils for public speaking competitions, for her it's important to monitor various performance factors. She supports the idea of practicing often outside the safe environment, because "practice makes perfect" (RM_TE).

Experts in the technical field agreed that audiences can be made to look like anyone know or unknown to the speaker and one of the simple ways is to use photos that can be then meshed on a virtual character's head.

The procedure of obtaining the facial features of the real person seems rather simple. One only needs a few photos that have to be edited and superimposed over a virtual character. In order to make this option functional within the virtual audience, the system should permit photo and video uploads and their meshing on the virtual characters (Table 3.4).

The situation becomes complicated though when the virtual character has to be animated to have the same facial expressions like the real person. If only static images are available, when animation begins, the virtual human might look eerie and nothing like the person it tries to embody. To help overcome this limitation, experts suggest recording people when they display various emotional expression, such as happiness or sadness, which then can be implemented in the animation. However, one expert points out another major difficulty of simulating known persons: behavior animation. Even if the face is created and animated in accordance with the traits of the real person, it is very difficult to grasp the personality of the person. Moreover, speech simulation complicates the matter as well, because the facial movements of the virtual human have to mirror the idiosyncratic movements of the person they copy (the way a person move their lips, the tics they have, etc.).

These technical difficulties suggest that a highly customized audience might only be advantageous for people who undergo long-lasting training courses and who prepare for targeted audiences (e.g., present in front of superiors whom they know and can help animate accordingly).

Table 3.4

How to design familiarly-looking virtual humans

Code	Interview excerpts	Conceptualization for virtual audience design
procedure	"[...] if you are using 3DMax [...] you just take a picture of a face you want to use and out it on the avatar face. That would be a very quick and easy process to do" (MNA_VRTechE); "I think in most cases you need a picture from like all sides of the head–two profiles and one frontal picture–to get a good 3D image of the person you want to do" (MNA_VRTechE).	photos/videos integration into the design and animation of VH
risks	"If I make you very nervous and you are somebody quiet, everybody will say that it's not you. [...] If you look generally happy, and I make you sad, it's not you" (DT_VRTechE).; "[...] if I take your photo, I can do your head, but when you smile, we do a smile which is not exactly you." (DT_VRTechE); "[...] if you want to make somebody speaking, there are people speaking fast, speaking slow, we have so many aspects that it is difficult to build a complete personality" (DT_VRTechE).	

3.4.1.1   Summary

The purpose of the inquiry on personalization was to see whether familiar audiences represent an opportunity for audience design. These results can be subsumed under the human-like attributes proposed in first level of the audience design concept.

Several studies showed that people felt more uncomfortable and performed more cautiously in front of familiar negative audiences than in front of complete strangers (Butler & Baumeister, 1998; MacIntyre & Thivierge, 1995). However, when familiar people are supportive, this has beneficial effects on performance (MacIntyre & Thivierge, 1995).

Experts in public speaking acknowledged the importance of varying audiences during training, by allowing trainees to practice first in a safe environment of known people, but also by encouraging them to practice more and more in front of unknown audiences. They recognized the importance of unknown audiences especially in preparation for competitions and for presentations outside the usual group of listeners (for instance to clients). However, they also admit that trainees feel comfortable when they address people they know, because it gives them a feeling of safety.

Familiarity can be further researched with regard to how someone is likely to behave in a given situation and the effects this could pose on speech performance. Previous studies showed that familiarity can lead to stereotyping, due to the fact that familiar objects receive less systematic processing (E. R. Smith et al., 2006). Future studies could look at whether this is the case when a speaker addresses an audience they know and what expectations they would have as well as how helpful they would perceive the feedback from people they know.

Whether personalization should or should not be an option in the virtual audience design is also something that trainees can decide upon. During training, it's good to have a supportive audience made of known people. However, performance improves outside of the comport zone, and this means actually not talking to familiar audiences. To solve this, a virtual audience could mitigate both needs and integrate both known and unknown virtual characters at the same time.

Technically speaking there are a few simplistic options to populate the virtual audience with familiar faces, such as meshing photos of faces on existing virtual humans, but with high risk of confusion and of reaching

the uncanny valley (Mori, 1970), if the virtual character looks familiar but behaves unexpectedly. A realistic option would be to use familiar faces in combination with impostor characters (Stuevel, Ennis, & Egges, 2014) who move little or sit in the back seats of the virtual room where they are less conspicuous.

3.4.2   Interaction

Interaction between speaker and audience can be discussed from the procedural angle of public speaking situations (interaction as feedback session at the end of the speech), the speaker-listener angle, and from the human-computer interaction perspective, as discussed in the theoretical chapter. The interview answers were first grouped along these subcategories of interaction. The research question was: *What is the importance of interaction between speaker and audience in the training process and how can interaction be acquired?*

3.4.2.1   Interaction definition

In order to respond to the research question on interaction importance in training and technical prerequisites to sustain it, the first step was to see what experts understand under interaction. For one of the public speaking experts (CS) interaction "means to react on something what you see in the audience, what you feel what comes from audience or you have feeling just for it.[...] There's no 100% right or wrong in that way as well". She also points out that interaction refers as well to how speakers react to the environment they occupy: "interaction also means to me, sounds a bit perhaps esoteric, but this thing how do I react to the room in which I am? How do I react to the atmosphere of the setting?". Speakers react to both audiences and the environment and both elements could influence speech performance.

Another training expert (RM_TE) points out the role of speech context in what interaction between speaker and audience looks like. This setting dictates levels of anxiety through factors such as place novelty and unknown audience members: in a competition setting "a speaker can be sidetracked or intimidated by a purposeful reaction from a member of the audience" (RM_TE). In the training setting, this kind of interaction is recreated, in order to prepare speakers to deal with potential similar situations, but in a safe environment. "The speakers have feelings (and they) are important to us. [...] and we try to

have interactions that are as constructive and friendly and supportive as possible" (RM_TE). Constructive feedback can be "highlighting something very good that the speaker should definitely keep in the speech" (RM_TE).

For two technology experts (TK, BH), interaction represents more the actions of the virtual audience: "the attitude in general it's a sort of interaction" (BH_VRTechE). Interaction is also seen as a bidirectional process through which both speaker and virtual audiences influence each other (TK_VRTechE). Interviewees further discuss two ways to achieve this, by either having someone control the reactions of the virtual audience, or by having an autonomous audience. Findings are presented below.

3.4.2.2   Speaker-listener interaction (during the speech)

One way to look at interaction is to consider the actors: speaker (one) - listeners (many). This type of interaction presented in Table 3.5 is represented by the codes "audience role", "speaker's role", "audience management strategies", and "question and answer sessions". Training experts' comments were biased slightly towards negative audience reactions, against which they also provide extensive management strategies.   Even if this might not be the case in real life, public speaking coaches prefer to prepare speakers for worst case scenarios. When audiences are truly disturbing, training experts provided a list of strategies the speakers could use to deal with them.

Trainers expect that speakers learn to evaluate audiences quickly and to adapt fast to audience signals, specifically if these are negative ones.   These techniques grouped under the subcategory "audience management strategies" give a glimpse of what speakers might have to deal with in a real speech situation. Among the mentioned strategies were eye contact, saying a joke, or making a noise, such as having a little bell nearby and ringing it, if necessary. Therefore, they are an indirect hint towards what behaviors virtual audiences could display.

Both speakers and listeners have clearly defined roles: the listeners have to react fast to speech content and style, whereas the speaker has to give the appropriate response: "it's very important that I react in the moment" (CS_TE). This constant exchange could be available also with a virtual audience and supported by fast adapting expressive

capabilities for verbal and nonverbal behaviors, in real time, to each new interactive situation.

Table 3.5
Speaker-listener interaction

Code	Interview excerpts	Conceptualization for virtual audience design
audience role	"people started talking in the audience [...] and people started laughing" (RM_TE); "it could be that the audience starts progressively to have a negative attitude if the person is not talking" (BH_VRTechE); "while the speaker is giving the presentation, somebody says 'hey, I don't like this' or somebody looks in a certain way" (RM_TE).	fast adaptation and expressive capabilities through verbal and nonverbal behaviors
speaker's role	"if I do my presentation and I've got some people who [...] perhaps disturb a bit [...] nonverbally, then I have to decide 'OK, what will I do'. [...] it's very important that I react in the moment" (CS_TE); "what can I do to include the audience, to make them listen to me, to help them to feel invited" (RM_TE); "try to make eye contact to 10-15 people in the audience" (RM_TE).	

Table 3.5
Speaker-listener interaction - continued

Code	Interview excerpts	Conceptualization for virtual audience design
audience management strategies	"eye contact [...] not to speak anymore" (CS_TE); "have a little bell [...] to change suddenly the sounds (in the audience)" (CS_TE); "ask 'can I help you? Am I not loud enough"' (TE); "always have a couple of anecdotes in your sleeve [...] chat them up" (RM_TE).	fast adaptation and expressive capabilities through verbal and nonverbal behaviors
question and answer sessions	"the farthest we've gone with simulating a hostile audience is to harass the speaker with questions at the end" (RM_TE); "if questions come from the audience [...] can be quite useful but it also can be quite annoying. Sometimes (the listener) wants to show he or she knows a lot and wants to offend you" (CS_TE).	

3.4.2.3 Interaction as feedback session (at the end of the speech)

From the procedural point of view, interaction refers to the feedback that audiences are encouraged to give to speakers at the end of their practice sessions (Levasseur et al., 2004; Lucas, 2012). This practice was discussed by one training expert who prepares pupils for public speaking competitions and was also briefly mentioned by one VR expert (BH). Interviewees explained that trainers practice feedback by preparing questions for the speaker and delivering them either in person or with the help of an avatar.

The codes for feedback are given in Table 3.6. First, feedback is meant to support and encourage the speaker (code "supportive atmosphere"): "we try to have interactions that are as constructive, and friendly, and supportive as possible" (RM_TE). Second, it is highly regulated, both in terms of duration and content (codes "duration" and "type of feedback"). One expert calls them "rules of giving feedback" (CS_TE). She also provides examples of how long feedback sessions last.

Trainers allocate extensive time for feedback at the end of each person's speech and they use these sessions also to teach speakers what they need to improve, as well as also teach them how to give positive feedback themselves (codes "benefits" and "type of feedback"). One of the most usual verbal feedback is questioning about the topic and discussing the characteristics of the speech, as one trainer affirms: "all members give them feedback and [...] tell them 'I think you got that introduction right"' (RM_TE). For public speaking competition, feedback on content is even more important than body language,

Table 3.6

Interaction as feedback in the training procedure

Code	Interview excerpts	Conceptualization for virtual audience design
supportive atmosphere	"we try to have interactions that are as constructive, and friendly, and supportive as possible" (RM_TE); "your feedback is supposed to help this person do a better job next time. So obviously they're (listeners) are nice" (RM_TE); "highlighting something very good that the speaker should definitely keep in the speech" (RM_TE).	valenced verbal feedback (preferably supportive); verbal feedback to speech content and presentation style; customization of feedback for each participant.
duration	"timewise it will be 20-25-half an hour for one person" (CS_TE).	customize feedback duration

Table 3.6

Interaction as feedback in the training procedure - continued

Code	Interview excerpts	Conceptualization for virtual audience design
benefits	"We have this situation of feedback from different people. (They) get used to the ideas 'OK, there are different people and here are different opinions as well"' (CS_TE); "we are sensible to the idea 'OK, what I see are nonverbal reactions from another (and) it doesn't mean that my interpretation is right"' (CS_TE); "collective feedback is very useful and productive at the beginning" (RM_TE); "say something that the speaker might consider changing, and make a suggestion" (RM_TE).	speech recognition, natural language understanding and production unit
type of feedback (on content or style)	"we look really carefully at [...] contents, because sometimes content [...] can be more offensive than body language" (RM_TE); "highlighting something very good that the speaker should definitely keep in the speech" (RM_TE).	speech recognition, natural language understanding and production unit; backstories.

because it has to be tailored to each audience in particular: "sometimes content [...] can be more offensive than body language" (RM_TE).

The conclusion that can be drawn here is that trainers prepare participants to face various opinionated responses from the audience either on speech content or body language. In a virtual reality setting, such opinionated responses can be achieved by endowing virtual

characters with background stories and with cognitive capabilities that support verbal feedback to a speech. Backstories are usually applied in gaming and, in a similar fashion like the audience characteristics in the speech effectiveness model (Verderber et al., 2008) they refer to the various demographic characteristics, cultural background, history or expectations of characters (Maldonado & Hayes-Roth, 2004; Rank & Petta, 2012). Several other elements that come along with backstory, such as identity appearance or manner of gesturing are presented in the chapter on the proposed training scenarios and virtual humans.

Interviewed experts mentioned also factors that can impact on the interaction between speaker and audience, some depending on audience members, some depending on speech location and setting: background of audiences, their political views, their preferences, the quality of air in a room, speech duration, etc (CS_TE, RM_TE). Also these resemble the audience characteristics proposed by (Verderber et al., 2008) and confirm the suggestions of the researchers with regard to audience diversity in speech preparation and delivery phases. However, such elements can also be used separately, to introduce the diversity of audiences and locations, before creating a complete scenario to train speech delivery. For instance, a virtual audience could be used to illustrate a certain age group, another could portray people with a common occupation, or a certain nationality.

3.4.2.4   Human-computer interaction

The last perspective on interaction is human-computer interaction. Experts in VR technology were required to discuss how interaction between user and audience can be technically achieved. The codes that occurred are "behavior recognition and devices to achieve interaction", "autonomy (no audience control needed)", and "synchronous control of audience" (control by a human operator). Interviewees mainly mentioned ways to create interaction and answers ranged from simple monitoring devices to complex ones. One expert pointed out that the user should not wear any devices in order to interact with the virtual audience: "in my opinion the user should not wear [...] even any joystick or so" (TK_VRTechE). Others suggested wearing unobtrusive devices, such as microphones in order to measure the impact of the virtual audience on the speaker and to transform this into input data for the virtual audience: "some devices measuring the information of the speech or (if the person) sounds unsure or not and then the audience can react

for example; you don't need any special hardware except microphone"
(TK_VRTechE); or use minimally intrusive devices, such as wrist
sensors that can help track a user in the virtual environment. Eye
gaze and speech volume proved to have an influence on how attentive
listeners become, with audience members paying more attention to the
speaker when the speaker had high eye contact and a normal voice, in
comparison to cases where the speaker used a low voice and looked at
the notes or at the projection of the presentation (Tudor, Mustatea,
Poeschl, & Doering, 2014). Hence, eye-tracking and voice volume
measurements would be realistic options for automated virtual audience
responses.

Another expert (RK_VRTechE) proposed more complex sensory de-
vices, such as those that measure galvanic skin response or heart rate.
He also grounded his preference in the fact that automated data record-
ing systems, such as senors can replace the presence of a psychologist
or of the person manipulating the virtual audience: "you can get phys-
iological measures that will inform the system of the state of mind or
the current behavior of the patient, you could have some automated
responses from the audience" (RK_VRTechE).

With regard to control of audiences, experts offered various solutions
and discussed their preferred functions that would enable behavior
control. Two types of simulation and audience control could be
identified in the interview answers: an autonomous type where the
audience responds on its own in real time to speech and speakers (code
"autonomy"), and a synchronous type where the trainer maneuvers the
behavior of virtual humans in real time, during the simulation (code
"synchronous control of audience").

The main findings on devices to achieve interaction, autonomy and
synchronous control are summarized in Table 3.7 .

Table 3.7

Behavior recognition and devices to achieve interaction

Code	Interview excerpts	Conceptualization for virtual audience design
behavior recognition and devices to achieve interaction	"you don't need any special hardware except microphone" (TK_VRTechE); "use Kinect" (TK_VRTechE); "head tracking and position of the wrist" (TK_VRTechE); "galvanic skin response" (RK_VRTechE); "heart rate monitor [...] or tracking the eye movement" (MNA_VRTechE); "feedback of [...] own voice, which really corresponds to the room acoustic situation in a real classroom" (TK_VRTechE).	language and body movement recognition components
autonomy (no control needed)	"listener [...] reactions should be determined by a set of (audience) factors [...]: age, profession, location, it would be really magic." (RM_TE); "you have a software and [...] this software tells the speaker if they did well or not" (RM_TE); "thought bubbles" (RM_TE); "psychological measures that [...] inform the system of the state of mind or current behavior of the patient [...] could have some automated responses from the audience" (RK_VRTechE).	speech recognition, natural language understanding and production unit; speech performance analysis unit.

Table 3.7

Behavior recognition and devices to achieve interaction - continued

Code	Interview excerpts	Conceptualization for virtual audience design
synchronous control of audience	"by pressing a button" (TK_VRTechE); "extra buttons for triggering [...] an extra event during the scene." (BH_VRTechE); "having like [...] a neutral setting, an interested setting, or a disinterested setting and the psychologist can actually just switch between settings" (MNA_VRTechE); "either you prepare thousands of millions of questions, either you do a text-to-speech, either you give him (the therapist) a microphone" (BH_VRTechE).	graphic user interface that allows trainers: (1) to control virtual humans within the same virtual audience and (2) switch between whole audiences (settings)

Both training and VR experts showed their support for synchronous control of audience behaviors. Through synchronous control trainers make use of predefined audience behaviors that they activate when needed: "I could control it, a little bit in that way that I see, OK, right at the moment I've got the feeling that he or she (the speaker) needs a bit more pressure from the audience" (CS_TE). The same expert also showed a lot of enthusiasm at the idea of manipulating a virtual audience in real time. Another one was however rather skeptical towards the success of this method, affirming that it could work, but "I don't know if it would work as a training tool for more than a few sessions" (RM_TE), since trainees would figure it out that responses are not real. Therefore, she would rather support an autonomous audience that not only reacts on its own to speakers, but also provides informative feedback on how well the speaker did. And, if verbal feedback is too complicated, virtual humans could express their thoughts through "thought bubbles" (RM_TE) that would pop

up above their head during the presentation, with brief information about their opinions on the speech.

Technically speaking, it is also possible to achieve synchronous manipulation of the virtual audience (code "synchronous control of audience"): "it is now programmed before, right, but then, just the selection of the actions can be performed by someone in real time" (RK_VRTechE). Otherwise one can also preset the virtual audience to run a certain type of scenario. Experts proposed a method similar to the one used in phobia treatment in VR: "the audience is reacting according to some transcript or the psychologist who is leading the experiment [...] can, by pressing a button, influence [...] the behavior of the audience" (BH_VRTechE).

Technology experts were skeptical with regard to autonomy and suggested ways to make control easy to trainers and therapists, since "the intelligence level of a human being is far deeper than what we can achieve currently by machines" (MNA_VRTechE). To achieve this control, trainers could use buttons to activate various audience behaviors and even switch completely between whole audiences to load new scenarios.

Automated audience reactions should not be completely discarded though, because there are many sensors that can pick up what speakers are doing, and the input can be used to analyze speaker behaviors in order to process an appropriate audience feedback. This would mean a three-steps approach: (1) recognize what the speaker is doing and saying, (2) analyze the data, and (3) process a feedback. Sometimes this led to the creation of own hardware: "[...] we have created special hardware, sensors to exactly measure what happens on the screen in order to synchronize these events perfectly with the reactions of the user" (TK_VRTechE). However, experts admitted that the procedure proved time consuming and costly.

3.4.2.5   Summary

The research question guiding the inquiries on interaction focused on the importance of interaction between speaker and audience in the training process and on how to acquire interaction between the two main actors in the virtual reality setting. Findings show that interaction is a very important topic in the field of public speaking and presentation training, with reference to interaction or aspects of

it in almost all interviews. Three types of interaction were presented: speaker-listener interaction during the speech, interaction as feedback session at the end of speech, and human-computer interaction, which addressed the technical conditions to achieve the first two types within a virtual setting. The presented results can be subsumed to the second level of the audience concept dedicated to interaction between speaker and audience, either via sensors or with the help of a graphic user interface manipulated by a trainer.

Feedback at the end of speech sessions and speaker-audience interaction during speeches are pervasive topics. Whereas nonverbal audience feedback is continuous during speech training sessions, trainers expect from seminar participants to give a well-thought and ample verbal feedback at the end of a trainee's speech. This can come in two forms: on the one hand, seminar participants who listened to a trainee are encouraged to tell the speaker what they think of the performance, on the other, they are also encouraged to ask questions related to the topic in order to simulate real life speech situations. The type of feedback (whether it's comments on speech performance or it's content-related questions) depends on the training format and training purpose. For competition training, feedback sessions are carefully monitored, especially with regard to the content (RM_TE). Speaking skills training outside the competition context is less of a rigid activity. Feedback is less structured and every participant is free to say what struck them most about the presenter and the speech. Such variety could also be reflected in the flexibility of a virtual audience where audiences are adaptable in their behavior to the needs of trainers and trainees.

Further on, experts in virtual reality technologies discussed the importance of having unobtrusive sensors that gather input data and can transmit the state of the speaker to a system running the virtual audience, which can, in turn, play certain responsive animations in real time. A brief review of literature revealed that when phobics experience anxiety, there are physiological responses to the stressor and an observable pattern of a speaker's behavior (Clevenger, as cited in Clevenger Jr., King, Clevenger, and King (1961)). The human body reacts to stress situations, such as public speaking, by secreting hormones, such as adrenaline, noradrenalin, or cortisol (Hennig, Netter, & Voigt, 2001). Measurement procedures include salivary cortisol monitoring before and after presentation (Garcia-Leal et al., 2005), heart

rate monitoring, skin conductance and breathing rate (Bassett, Marshall, & Spillane, 1987; Westenberg et al., 2009). Each of these physiological measures have specific procedures and timing of measurement. Whereas skin conductance, heart rate and breathing rate can be measured continuously during the speech procedure (Westenberg et al., 2009), saliva has to be collected from the user before or before and after the speech task (Bassett et al., 1987).

For a real-time interaction between speaker and virtual audience to occur, it's important to consider the processing time of recorded physiological data and the way it can be fed into the system to trigger virtual audience reactions. Moreover, in the studies mentioned above, subjects were also required to avoid any caffeinated drinks, cigarettes, alcohol, or making any intense physical exercises prior to the measurements (Westenberg et al., 2009). To implement such physiological measure in a usual training procedure would require dietary changes that could complicate the procedure unnecessarily. Apart from that, all these measurements were used in experimental settings, and not as real-time feedback in an ongoing interaction between speaker and audience in a speech training procedure. Future studies could inquire whether anxiety levels of non-phobic speakers can yield relevant data in order to justify the use of sensors, as well as identify the most fitting ones. One could also compare sensors with different degrees of invasiveness and asks trainers and trainees about the method they would prefer most.

The same would apply to using speaker's nonverbal behaviors as input for interactive autonomous audiences. Clevenger Jr. et al. (1961) mention several behaviors associated with stage fright which are also easily visible, such as feet shuffling, swaying, arm swinging, lack of eye contact, hands in pockets, back and forth pacing, or playing with objects in their hand. Such behaviors are trackable with the help of existing sensors, such as Kinect, as one expert (TK_VRTechE) already mentioned. This method would also have to be tested to see whether non-phobic people provide enough information about their inner state so that virtual audiences can respond accordingly, for instance by being attentive or bored. More advanced methods of interaction would involve natural speech processing methods, such as those used in ECA design (Cassell, 2001), but VR experts seemed rather skeptical about the ease of creating such complex systems. Methods like these could be used more for the feedback sessions at the end of speeches, where virtual

audiences could ask content-related questions. One has to consider though the importance of such sophisticaled audience capabilities for more advanced speakers, where successful Question & Answer (Q&A) sessions are very important. For less advanced speakers, it may be sufficient to have an audience that reacts only nonverbally. Based on these findings, one could consider the option of devising several training packages that contain more or less autonomous audiences with various speech capabilities (i.e., no dialogue, some, and extensive dialogue).

3.4.3 Group dynamic

In the theoretical chapter of the study, group dynamic was referred to as representing those behaviors that occur within real audiences such as applauding or shifting attention towards a stimulus in a ripple effect that engulfs more and more people. In a virtual reality setting, group dynamic was referred to the same effects and also to making groups react unitary, while keeping their behavior diversity. The research question was: *What group dynamic behaviors occur within a virtual audience and how can they be implemented in a virtual audience?* It inquired the typology of group dynamics that occur in real life speaking situations and the technical possibilities to implement them.

Public speaking experts were required to discuss whether they noticed any type of behaviors that engage the whole audience and describe their characteristics (code "behavior type"). Apart from behavior types, experts discussed about characteristics such as "trigger" and "spreading pattern". Findings are summarized in Table 3.8.

One expert stated that it is usual for pupils to start engaging with their colleagues and this is a behavior that spreads very quickly and is very hard to control. Such behaviors are relevant for large audiences, like the one devised by North et al. (1998), because these offer more playroom to unfold. Triggering factors could be someone outside the virtual room's window. Sometimes the speaker's voice is too low and people start asking questions. Other times people simply distract each other by throwing things across rows and disturbing fellow students.

Table 3.8
Group dynamic behaviors

Code	Interview excerpts	Conceptualization for virtual audience design
behavior type	"if people can't hear you [...] they start to (ask each other) '[...] do you hear her?"' (CS_TE); "chitchat the whole time" (RM_TE); "someone didn't understand anything [...] and she is asking the neighbour [...] and it goes further and further" (CS_TE); "laughing"(RM_TE); "interaction over rows, like [...] throwing things" (RM_TE); "something happens outside [...] one watches, the next one watches and then everyone watches" (CS_TE).	interaction with virtual environment (e.g., manipulating virtual objects) and other virtual humans
trigger	"something happens outside" (CS_TE); "a neighbour" (CS_TE); "you could have random targets for the audience to look at" (RK_VRTechE).	interaction with virtual environment and other virtual humans
spreading pattern	"asking the neighbour [...] and it goes further and further" (CS_TE); "interaction over rows" (CS_TE); "something happens outside [...] one watches, the next one watches and then everyone watches" (CS_TE); "having some final state-machines that would [...] define this group dynamics" (RK_VRTechE).	interaction with virtual environment and other virtual humans

An expert expressed his opinion on virtual characters that behave and react independently to a speaker: "[...] I will say that each person in the audience, more or less tries to behave as an individual [...] unless the audience is really really mean [...] usually they won't behave in a group" (MNA_VRTechE). Another one confirmed what has been discussed in literature (Ulicny & Thalmann, 2002) and stated that technically it is possible to obtain a group reaction through mutual monitoring by having random targets for the audience to look at (RK_VRTechE). In order to achieve this, the system should permit the virtual humans to interact both with objects and other virtual humans within the virtual environment.

3.4.3.1 Summary

The inquiry on group dynamic was meant to present a behavior typology of a real audience and to introduce some technical aspects about how such behaviors can be achieved. Public speaking experts mentioned several cases of audience unrest they've experienced and how the behaviors spread, as well as what triggered them. Experts in design and programing confirmed that virtual audiences can be created to react to both speaker and other fellow audience members. Although they didn't suggest exact procedures on how to do so, the literature on interaction within groups (Benford & Fahlén, 1993) can serve as a starting point to achieve dynamic audiences.

3.4.4 Real life training

The forth inquired topic was the actual speech practice part during a training session. The addressed research question was: *How does a training session look like and how does it accommodate the speech practice and speech feedback?* The reported findings were obtained through the observation of the speech practice session as well as through the inductive content analysis of the interview transcripts (category "training procedure"). During the discussions with the public speaking and presentation experts and with some of the technology experts, a large amount of information was gathered on how training procedures are organized. The observation complemented the data with further details about the way a training seminar is conducted. All results are pooled together in the tables and the ones from the observation are marked with an "O".

Understanding training procedure helps structure practice sessions with the virtual audience–such as deciding when to use a virtual audience and for what kind of learning content. Understanding how training sessions are conducted and what their main components are helps organize design–what demographic features have to be considered for the audience members and speaker, what is the scenario setting, and how long should a scenario be. The information provided here gives a glimpse into the third level of the audience concept that deals with the organizational aspects of communication training and the role a virtual audience might occupy in its context.

Table 3.9 summarizes some of the general elements of training procedure: who are the trainees, what is the speech format they practice and train for, as well as speech duration.

The pool of potential trainees is very vast. Age, occupation, country of origin, education varies greatly and this reflects both the characteristics of the speaker, but also of the training group it belongs to.

For instance, one trainer (RM) recalled preparing a group of employees. They all belonged to the same company, to the same department and the audience for each speaker was made of their fellow colleagues. Such groups are rather homogeneous because trainers are hired by institutions to offer a targeted training seminars (RM_TE, CS_TE). Therefore, everyone who is a speaker will share the characteristics of their fellow trainees, such as membership to a similar group (students, employees) or country of origin.

Age, occupation, country of origin, education can be varied within a virtual audience to either match the characteristics of the speaker, as in a homogeneous training seminar, or to offer a totally different audience experience: for instance have a German student speak in front of a foreign audience at a university abroad; prepare a scientist explain complex data to potential clients or investors. This audience variation meets the desiderata of communication in the disciplines and communication across the curriculum paradigms (Dannels, 2001, 2002), which encourage people to become competent speakers independently of their field of work and study.

Other audience characteristics variations can come indirectly from a speaker's reason to join. As speech trainers mentioned, some people wish to improve self-confidence, some wish to become leaders in their

Table 3.9
Training procedure - general features

Code	Interview excerpts	Conceptualization for virtual audience design
trainees	"12-18 in school, and [...] in the business sector 20, 20 something to [...] mid-40" (RM_TE); "mother tongue German [...] from 20-60" (CS_TE); "students from totally different countries [...] like engineering students" (CS_TE); "sometimes scientists" (CS_TE); "at least 25 nationalities" (CS_TE); "I also train for the National Debate Association, [...] companies" (RM_TE).	various demographic characteristics of virtual listeners and outfits
speech format	"professional speeches [...] like in corporations, competition (speeches)" (RM_TE).	audience customization to match speech format and speech context (e.g., a business gathering)
duration	"3-8 minutes" (TE); "one minute formats" (TE).	duration of virtual scenario

groups, others wish to improve their presentation skills in the last years of school, improve their communication skills as scientist, or learn specific skills, such as preparing presentations or managing Q&A sessions. This further supports the need for artificial intelligence that would reflect in a virtual character's cognitive and affective capabilities.

The need of high audience customization is both an advantage and a disadvantage for a virtual audience. High customization can help train a large group of potential users, but also result into pressure on the design process, by requiring great flexibility of audience looks, behaviors, and of speech scenarios.

3.4.4.1   Speech preparation

Before getting to the speech practice session, there is an ample speech preparation phase. One expert synthesized the two basic elements of training (RM_TE): "what you say and how you say it".

Both presentation content and style are discussed by the trainer and the next point is to see the role of a virtual audience in this process and how it could help trainees.

Speakers are required to provide information about the context where they will have to perform, in order to help trainers plan customized solutions (code "speech preparation phase" in Table 3.10). They also have to learn audience management methods, which have already been discussed in the subchapter above on speaker-listener interaction. Table 3.10 summarizes what speakers have to consider before starting to practice in the training session or giving the speech outside the session. Trainers help them figure out the audiences they prepare for, the hierarchies within them, as well as the speech structure. Mainly it can be summed up as "what should I do to fulfill my own wishes as an audience" (CS_TE).

Sometimes speakers don't know anything about the audiences they will address in real life. If this is the case, the trainer is the one who provides the information about the audience and speech context for which they should prepare.

The speech preparation phase is relevant for the initial virtual audience customization phase. In this phase, trainees have the opportunity to think of their potential listeners in the real world and have them represented as precisely as possible in the virtual world.

Table 3.10

Training procedure - speech preparation

Code	Interview excerpts	Conceptualization for virtual audience design
speech preparation phase	"asking before [...] the presentation what is the room, who are the people, what's the time" (CS_TE); "[think about] who am I, what is my position for or in connection to the audience" (CS_TE); "(think about) different roles, for example you have an audience which includes [...] professors and simple students, both parties have totally different backgrounds" (CS_TE); "look at different countries, depending on who we expect to be in the audience (at the public speaking competition)" (RM_TE).	various demographic characteristics of virtual listeners, outfits, and spaces they can inhabit
duration of speech preparation	45-60 minutes (Observation)	customize scenario duration; adapt behavior duration to the overall scenario timing.

3.4.4.2 Speech practice session

The five-hour speech training session observed as part of a three-days communication training seminar at TU Ilmenau helps further with putting the speech practice into the larger context of the training program (the third level of the virtual audience concept).

Table 3.11
Speech practice seminar structure

Seminar day		Observation excerpts - activities	Conceptualization for virtual audience design
day 3	theoretical part	teaching skills, such as: style of language, style of speaking; learn performance criteria: such as organizing presentation slides, using certain gestures, importance of free speech, reacting and interacting with audiences (approx. 3 hours).	speech recognition, natural language understanding and production unit.
	practical part	a non-compulsory speech practice session of 5 minutes per participant; thorough feedback session of 45-60 minutes.	various demographic characteristics of virtual listeners and outfits.

Compared to the total time allocated to skills teaching, the practice part itself, although very relevant in the opinion of the trainer who had been interviewed beforehand (CS) covered only the second half of the last seminar day. The speech practice session of this particular seminar was much shorter than both the theoretical part preceding it and the feedback session that followed, with the feedback session stretching up to one hour. An explanation for this could be of educational nature: it is important to first acquire theoretical knowledge about speaking skills and then practice them.

Each participant was invited to give a presentation about themselves and was given five minutes time for the speech (See Table 3.11). The trainee was invited to go to the front of the classroom and was

allowed to use flip charts and any other visual aids they wished. Presentation time was strongly enforced and people who took longer than five minutes were interrupted by the trainer after another minute. After each speech, participants simulated a short Q&A session and the speaker was given extensive feedback.

3.4.4.3   Role-play

Trainers require speakers to imagine audiences they would address in real life or describe the ones in front of whom they will have to present, in order to make them visualize the speech context. Then, during the training session, these would have to imagine speaking to the audience they had in mind (CS_TE). Imagination is a skill that requires training, as one expert affirmed: "so like the sportsmen do it before, this really mental training. But if you are not used to mental training, it's a bit tricky" (CS_TE).

However, instead of speaking in front of an imagined audience, fellow trainees could role-play one. Experts were asked what they thought about having trainees in the audience to role-play certain audience types to help the trainee enter the speaker role better. Experts were rather ambivalent about using role-playing techniques during training. At first uncertain about telling trainees to fake their behaviors and be extra nice or tough towards the speaker, they didn't exclude this practice completely: "It's a good idea!" (CS_TE).

Even if it didn't seem like a common practice for the interviewed trainers, role-playing took place once in a while. Table 3.12 summarizes the results.

One trainer proposed a method to use role-play to measure performance with the help of a virtual audience, by establishing rules of speech and then testing speakers (RM). She suggested teaching several skills (e.g., do not interrupt when someone is talking) and then using the sensorial and feedback capabilities of a virtual audience to test whether speakers respect the rules or break them, for instance in a feedback session when verbal exchanges between speaker and listeners are less structured. This could be a systematic way of learning skills, by practicing them in different audience scenarios and having them evaluated within a concrete scenario.

The virtual audience could ease and speed up the training process by allowing speakers to train with various audiences instead of performing

energy-consuming role-play during speech practice session. Nevertheless, in order to be effective and offer a useful alternative to role-play, virtual audiences would have to be accessed and set up fast, to save as much time as possible for the speech practice itself.

Table 3.12

Training procedure - role-play practice

Code	Interview excerpts	Conceptualization for virtual audience design
implemented behaviors (have already been used by the trainers)	"make a noise halfway through the speech that would disturb the speaker (RM_TE); "really not encouraging audience, people who really want to get you, and ask you uncomfortable questions" (RM_TE); "harass the speaker with questions at the end" (RM_TE).	expressive capabilities through verbal and nonverbal behaviors
wish list behaviors	"say 'I don't like this'" (RM_TE); "have some ground rules [...] 'do not interrupt' or 'do not do this or that' and [...] simulate a difficult situation for a speaker and (the virtual audience) says 'you break rule x, you break rule y you break rule z during the speaker presentation'" (RM_TE).	expressive capabilities through verbal and nonverbal behaviors
reasons to be skeptical about role-playing	"the seminars [...] are not long enough to develop a real big presentation" (CS_TE); "these people are colleagues in the same public speaking club and they kind of tend to make it funny" (RM_TE).	virtual audience that are easy to set up and can be accessed fast.

3.4.4.4   Feedback sessions

One of the highlights of the training procedure is the feedback session at the end of the presentation session, which includes the feedback from other participants and the feedback of the instructor. Table 3.13 summarizes the findings of the training session observation. These findings complement the results presented in the interaction chapter and highlight some procedural aspects of how feedback is organized (codes "feedback qualities" and "trainer role").

In the feedback part, participants were required to pay attention to the speech in a global sense, without particular focus on specific verbal or nonverbal elements that should be improved. Therefore feedback seems to have little structure (code "feedback qualities"). Both the trainees and the trainer gave tips on various body movements and postures that came into discussion (code "feedback type") in a random fashion and every participant could say what they *recalled* about the speech and the content. A form of more standardized speech assessment was used by (Hughes & Large, 1993) who reports of a study where fellow students and teachers assessed speech performance of other students. Interestingly, although students were tougher in grading their peers, both groups awarded all in all similar grades to the speakers.

A virtual audience could be used to offer *simultaneous* feedback during a speech practice and embody specific listener behaviors. For example, by using microphones, eye tracking and motion sensors, the virtual audience could respond in real time (act bored or attentive) to a speaker, depending on trainee's voice volume or gaze patterns.

Another suggestion is to allow the other training participants to assist when the speaker presents in front of the virtual audience. Making the virtual audience available to everyone during speech training practice simulation could help all participants imagine the speech situation as well as see how audience reactions are connected to what the speaker did or did not do well.

It is commonly agreed in the communication skills field that public speaking training contains an instruction phase of skills, a practice phase where skills can be applied, and a feedback phase (e.g., (Docan-Morgan, 2009; Hargie, 2006; Green, 2003)). Several of the reviewed studies in speech training and treatment focused only on the speech practice phase. However, due to their attention on speech practice

Table 3.13

Feedback round in the observed training session

Theme	Observation excerpt	Conceptualization for virtual audience design
feedback type	feedback on presentation style and content	expressive capabilities through verbal and nonverbal behaviors in order to offer targeted and structured feedback to speech content and style
feedback qualities	more of a free discourse; similarities to a brainstorming session	
trainer role	last one to give feedback; more thorough; gave examples of good speech delivery practices; synthesized the take-home message	
drawbacks	little relevant feedback on content, due to the presentation topic ("talk about yourselves")	

impact on anxiety levels, neither the processes that take place during a speech practice session nor the role of the trainer/therapist are explained in detail. The findings presented here are meant to fill this gap and shed light on some of the actual practices that take place during training.

3.4.4.5 Summary

Training experts shared how they organize their seminars and revealed the importance of steps such as the theoretical preparation for the speech, speech practice, and the extensive feedback sessions. These sessions are highly valued and trainees are encouraged to give feedback to the speakers both on content and on presentation style. Interaction between speaker and listeners during a speech, such as role-playing an audience who is not interested in the speech occurs as well. However, this is less encouraged than the systematic feedback at the end of the presentations. When asked about their willingness to use a virtual audience to embody the natural diversity of a public, trainers showed

interest and enumerated all kinds of behaviors they would like the virtual audience to display, from the very supportive, to the very critical ones.

The expectations of trainers from a virtual audience are high and comparable to the ones they would have from a real one: understanding of presentation content, display of diverse behaviors, and also the possibility to control these elements easily.

3.4.5 Virtual audience characteristics

Some of the most extensive results belong to a category that emerged through axial coding: "general application characteristics". Many themes could be gathered here and they cover information about virtual humans' behavioral and physical characteristics described by experts, as well as what experts expect from virtual audiences or what they consider risky in the design processes. Virtual audience characteristics represents a subcategory of the "general application characteristics" and contains relevant information about (1) physical characteristics of virtual humans, (2) listener behaviors, and (3) specific feedback behaviors between listeners and speaker considered necessary by the interviewees. This chapter brings new information about the human-like attributes expected from virtual characters, as proposed by the first level of the audience concept. Moreover, it also exemplifies how listeners behave in real speech settings, therefore giving concrete suggestions on how to achieve autonomous, interactive and personified virtual humans (technical attributes in the first level of the audience concept).

Before delving into precise behavior examples, one can make use of a simple audience categorization proposed by one of the experts (RM_TE):

- the well-behaved audience which is neither bad nor good, is polite, and keeps good eye contact with the audience;

- the enthusiastic audience where people are visibly involved in the speech, where people "light up at what you're saying";

- the "oh, come on" audience with people who sometimes are forced by circumstances to attend a speech.

This simplistic typology anticipates the audience valence (positive and negative) and also some of the behaviors that will be discussed next.

3.4.5.1   Demographic characteristics

Experts enumerated specific physical characteristics that virtual humans could embody. These are tightly related to the experts' practical experience and to their targeted audiences (e.g., students who speak in front of same-aged people, speakers from other cultures than the listeners, etc.). Along with these characteristics, experts mentioned also speech settings, therefore it was decided to couple these categories in Table 3.14.

Training experts mentioned many demographic characteristics that can be further divided in two typologies:

1. those that can be displayed, such as age, facial features, race, and outfits (codes "general demographic components", "physical characteristics" and "outfit") and

2. those that can be embedded in a virtual character's personality, such as their preferences, hobbies, or beliefs (code "general demographic components").

An important finding from the discussions about these characteristics with the training experts was a phrase they used often "I would like to be able to choose", which denotes that they would expect an audience that can be customized on several dimensions. For now, one can propose the dimensions identified here as starting point for a user interface dedicated to trainers: general demographic components, physical characteristics, outfit, status, audience size, and virtual location.

Status for instance describes here the usual groups speakers train with. These are normally peers (school or work colleagues). Sometimes though they have to perform in front of jurors (public speaking competitions), managers, teachers, or professors. Status though is not a feature that can be easily inferred from outfit or age, but rather from the speech situation. Therefore, one solution to successfully implement virtual humans with a different professional status than the speaker would be to frame the speech as such from the beginning.

Table 3.14

Demographic characteristics of virtual audiences and their location

Code	Interview excerpts	Conceptualization for virtual audience design
general demographic components	"the person in the audience could have same sex, race, religion (as the speaker)" (RM_TE); "I could choose the profession, [...] some of their hobbies" (RM_TE); "something about likes and dislikes [...] or country" (RM_TE).	audience customization options via database and graphic user interface
physical characteristics	"long hair" (CS_TE); "beautiful girls [...] and beautiful men as well" (CS_TE).	
outfit	people dressed really formal [...] black suits" (CS_TE); "some hats, some caps [...] because lots of students wear it, especially some foreign students" (CS_TE); "casual outfit" (CS_TE).	
status	"colleagues in the same public speaking club" (RM_TE) "(audiences) at public speaking competitions" (RM_TE)	
audience size	"a scenario with one, with two, with ten, with 20, with 1000" (BH_VRTechE).	customization options for audience size, location size, and place assignment
virtual locations	"in an open space or in a school classroom, or in a bar" (RM_TE).	

Furthermore, it becomes clear that the audience characteristics given here might require specific technical considerations: various audience sizes should be coupled with various locations that can accommodate the chosen audience as well as with randomization rules that impede unnatural repetitions of looks and actions.

3.4.5.2   Virtual listener behaviors

Listener behaviors were divided into two types: virtual audience independent behaviors and virtual audience feedback. The first type refers to listener behaviors that do not explicitly depict a response (feedback) towards the speaker. Of course, any audience behaviors can be interpreted as reactions to what the speaker is saying, as is the case in human interaction (Poyatos, 1985), particularly if these reactions can be traced back to what a speaker was doing when they occurred. The activities reported here don't account for what speakers were doing in the speech practice or should do in the future to trigger them and ought to show behaviors in a "decontextualized" manner. The second type describes audience behaviors influenced by the speaker.

Listener-independent behaviors of virtual audiences

The data on virtual listeners behavior was initially very heterogeneous and contained a lot of scattered, yet detailed information recalled by training experts. After several recoding procedures, it was possible to group it along two features: (1) Body modality–verbal or nonverbal and (2) Valence–positive, negative, or neither nor. Example of this kind of behaviors are presented in Table 3.15.

Most of the discussed listener behaviors had a negative connotation and involved mostly nonverbal manifestations. The verbal behaviors that occurred could also be assigned under a negative valence, because they involved listeners talking to each other. This could be further associated with the group dynamic category, as example of behavior that can start with a few people and then lead to a generalized state of restlessness.

Table 3.15
Virtual listener behaviors

Valence	Subcategory of body modalities	
	Verbal	Nonverbal
positive		"has [...] the face in the direction of the speaker [...] changing posture but not moving much, [...] attentive" (BH_VRTechE)
negative	"virtual humans in the audience talk to each other" (TK_VRTechE) "people speaking in the background" (TK_VRTechE)	"you've got the feeling they judge you, they write something down" (CS_TE); "someone comes later and he has to go to a chair in the middle of the row [...] it's this murmur" (CS_TE); "people yawning" (CS_TE); "have a phone ring" (RM_TE); "looking back again and again, [...] at the window (CS_TE); "playing on the mobile phone" (CS_TE).
neither nor		"drinking, eating" (CS_TE); "people laughing" (CS_TE).

Listener-dependent behaviors of virtual humans

Table 3.16 represents a type of behavior where concrete listener behaviors trigger distinctive audience reactions. The conduct described here presents behaviors as feedback towards the speaker. The role of the speaker can be deduced from the interview segments.

Compared to the independent listener behaviors described above, the feedback behaviors are strongly targeted towards immediate reactions of the virtual audience. In order to implement such behaviors into a virtual audience, one would have to consider the cognitive and emotional characteristics of the virtual humans and to find a way to

either have them controlled by trainers or to achieve actual autonomous agents that understand speech content and presentation style, as discussed in the chapter on interaction.

None of the training experts had previous experience with virtual reality, therefore the topic of behavior realism was not discussed with them. Behavior realism appeared though in the interviews with technology experts, who give several suggestions on how to achieve it. There were two rules of thumb: have a virtual human that "doesn't look like he doesn't understand" and try to avoid the uncanny valley. Realism was translated into having the virtual humans look in the direction of the speaker, move smoothly like a real human would, and make them look differently from one another. One expert actually even argued against too much realism, by warning that "the more realistic a character looks, the more it has to act exactly like a human being. That is why people react better to a cartoon-like character than (to) a character that looks closer to a real life person" (MNA_VRTechE).

Table 3.16
Virtual audience feedback

Valence	Subcategory of body modalities	
	Verbal	Nonverbal
positive	"some kind of verbal interaction like 'you are late, you are over the time' and that can be said quite friendly" (CS_TE); "people say 'yeah, OK"' (CS_TE).	"people who look very friendly" (CS_TE); "having them clap" (RM_TE); "nodding" (CS_TE); "smiling back" (CS_TE).
negative	"just interrupt" (RM_TE); "disagreement [...] really loud" (CS_TE).	"boo, stand up and leave the room" (RM_TE); "kind of annoyed 'neaaah' [...] contempt" (CS_TE).
neither nor	"have the audience ask questions" (RK_VRTechE); "there should be a verbal feedback, because we get it too less" (CS_TE).	"interjections like 'aha, 'hmmmmm?"' (CS_TE)

3.4.5.3   Variety

Variety was inquired together with group dynamic but was coded under general audience characteristics, because it's also related to realistic looks and behaviors. Variety can manifest through creating diverse physical features and different behaviors and by randomly assigning such features to audience members. In order to avoid repetitions, audience behaviors can occur randomly. On the one hand, there is the bottom-up approach (code "achieving variety bottom-up"), on the other, there is the top-down approach (code "achieving variety top-

down"). In general terms, variety can be achieved by randomizing audience demographic characteristics, their behaviors and behavior occurrence, order of occurrence, duration, and intensity. How many behaviors can a person actually display, how much they vary from another one, timing, and order of occurrence are topics discussed thoroughly in the second study of the dissertation. The findings of the behavior observation are clear examples of how behavior timing and order can be varied, as well as also which behaviors can be varied, and which are relevant for obtaining a diverse and non-repetitive virtual audience.

Randomization is used as a technical term that helps an observer determine how much variety is depicted and also helps establish how much behavior variety is necessary to keep the simulation realistic (see Table 3.17). One virtual reality expert used a top-down approach to variety by devising general demographic characteristics: "I want to have 100 people, I want to have let's say 40% men and 60% women, I want to have 30% in row 20, 20% between 20 and 30 and so on [...] Or I give for example [...] the [...] size of the people, so the people can be generally tall or they can be short. I will not decide for each individual" (DT_VRTechE).

Interviewed training experts felt closer to the bottom-up approach, speaking of individuals (e.g. RM: "three of them are Germans, one is French, and one is Gypsy"). In the design process, such view differences could lead to different priorities of selecting behaviors and controlling them via the graphic user interface.

If trainers wish to have an audience with a majority of attentive behaviors, they could use percentages to describe what the audience should do: percentage for how many attentive and how many inattentive, percentage for how much of the time they should act in a way or another. Then, the system would have to randomly assign behaviors to each virtual human, based on these percentages. If trainers wish to decide for each person in particular how these should behave, they would need a graphic interface that allows them to select each person and assign them the desired characteristics. Further research should help clarify which would be the easiest method to put into practice these requirements.

Table 3.17
Achieving variety in virtual reality

Code	Interview excerpts	Categorization for virtual audience design
achieving variety (bottom-up)	"model four, five different actors [...], and then randomize them into five or six hundred [...] it can be done by a machine" (MNA_VRTechE).	expressive capabilities through verbal and nonverbal behaviors; first assign *each person*, than randomize
achieving variety (top-down)	"You will not generally decide a specific person. We will [...] decide the general audience." (MNA_VRTechE); "the average way [...] the people can be generally tall or they can be short." (DT_VRTechE).	expressive capabilities through verbal and nonverbal behaviors; first assign characteristics and behaviors to the *whole group*, than change proportions
randomi-zation categories	"randomization of the appearance, [...] body language, or [...] facial expressions" (MNA_VRTechE); "trigger them (behaviors) at different times [...] or different orders" (RK_VRTechE); "display that interest (in a lecture) in a [...] slightly different fashion" (MNA_VRTechE).	expressive capabilities through verbal and nonverbal behaviors; system and interface options for: type of behavior, time of occurrence, order of appearance, intensity of behaviors, idiosyncratic display.

3.4.5.4   Summary

The findings presented in this subchapter show the many ways one can describe audience characteristics and their behaviors. This is a proof of the natural group diversity that can occur among listeners. The issue of diversity becomes relevant once the number of virtual humans depicted raises, therefore the solutions enumerated above can be added to the first level of the virtual audience design, which entails all virtual human characteristics. Experts described and imagined audiences that vary along many demographic characteristics, various outfits, and behaviors. Trainers stated that audiences are rarely totally homogeneous, and, having such a large choice of audience characteristics would be in their view an extraordinary opportunity. However, this variety can be advantageous for big groups of trainees who are themselves heterogeneous and require customized training, whereas for homogeneous groups (e.g., employees in the same department in a company), a system with high customizable audiences could prove unnecessarily complicated. Feasibility of such highly customizable audiences could be approached in future studies.

Having so many potential characteristics and behaviors can become problematic and prove difficult to manage. Therefore, experts were required to discuss ways to achieve such variety, without jeopardizing audience realism. Due to the general approach of the presented study, it was unlikely that highly technical discussion would take place. For this reason the information rather confirms existing themes in literature than brings unprecedented data. A difference also emerged between two potential ways to customize audiences, either in a top-down or a bottom-up approach. Assigning behaviors by chance can be a successful model for large crowds. However, trainers expect certain behaviors to occur with precision, at certain points during the speech and as an explicit reaction to how speakers deliver the presentation. These aspects can be further researched and used to lay a common ground for both procedures in the design of the user interface.

3.4.6   Virtual audience setup

The topic covered here summarizes some of the findings on technical data. The findings were identified through inductive coding and were brought into discussion both by trainers and technology experts. They are complementary to the results on speaker-audience interaction and present solutions to the audience setup process before the begin of a speech. In order to interpret the findings, they have to be put into context and explained through the eyes of the trainers and therapists.

A trainer's needs with regard to the virtual audiences can be summed up as: high audience customization and simplicity of control. As one technology expert affirmed (BH) "they (trainers and therapists) don't want to mess with the technology". Coaches expect to be able not only to set up an audience, but also intervene fast in the simulation if needed. This applies to speech practice sessions in two concrete situations: (1) the trainer controls how the audience behaves during the speech and (2) the trainer uses the virtual audience to explain and illustrate demographic and behavior audience typologies.

Both high customization and simplicity of control are desiderata and, for a better understanding of what they mean in concrete ways, experts would have to become part of the product development process through usability testing (Redish, 2007).

Experts discussed ways to set up a virtual audience before a speech and then control it during the speech through an interface. Audience control was also approached in the human-computer interaction subchapter in the context of audience autonomy levels. It was revealed there that experts favored audiences they could control in real time during the speech and they'd like to chose between various audience behaviors. Here, concrete GUI elements are given in Table 3.18.

It is expected that audience characteristics are easily selected before the simulation starts and, once this begins, VA behaviors can be controlled. The interface should contain buttons, drop lists, and options to save trainees' performance in order to monitor skill progress. One public speaking expert (RM_TE) used the term "profile" to name different audience setups with various characteristics that can be used in the training.

Table 3.18
Audience setup

Code	Interview excerpts	Conceptualization for virtual audience design
graphic user interface	"the therapist should also have a history of what he did so that he can make progressive increase in difficulty" (BH_VRTechE); "a drop list" (CS_TE); "it would be the same as any computer game really, [...] you have some settings [...] at the beginning and then you just play" (RM_TE); "setting things like audience size, [...] how many men, how many women, what age, what place, [...] be able to chose the profession, [...] personalize likes and dislikes, [...] personal hobbies" (RM_TE); "a screen that shows [...] a little menu for each of the virtual human in the audience" (RK_VRTechE); "a set of buttons" (BH_VRTechE).	graphic user interface

As shown in Table 3.18, trainers expect that the interface would be able to accommodate virtual audience members demographic, behavioral characteristics, virtual location, and audience size (as discussed in the virtual audience characteristics section above), as well as duration of simulation. The interface can therefore be used in three ways: (1) to set up the audience with its characteristics, behaviors, and randomization rules, before the speech; (2) to control the audience in real time, during the speech, as discussed in the chapter on interaction; and (3) to save users' profiles with their customized audiences for training monitoring. Public speaking trainers showed great enthusiasm at the idea of being able to control virtual people and have them play a certain behavior in order to test the speaker and to help improve speech skills. The

suggestions they made with regard to the elements of the interface can be researched in depth in future studies and a GUI could be developed together with and tested by trainers in a real speech training situation.

3.4.7   Limitations and risks in virtual audience design

The interviews provided rich and complex information about training procedures and also a lot of high expectations on behalf of trainers. This last section represents data coded inductively with regard to technical limitations and risks in the design procedure.

Whereas the two public speaking trainers were very enthusiastic about the idea of a virtual audience, technology experts were more cautions in their approach. They helped highlight also the complicated side of virtual audiences, which contains several limitations and risks. Main limitations stem from technical capabilities with regard to scenario realism and virtual human interactivity. For instance, realistic familiar faces can be achieved only if animations behave similarly to their real counterparts (as discussed above in the section on personalization). Also, if the virtual humans behave in an extreme fashion, they can land in the uncanny valley or become visually unfit, which could scare off participants. Even if trainers would be happy to have listeners that look familiar to the speaker, the design and animation processes seem tedious and risky. Another risk comes from cybersickness, specific to 3D environments (people could feel very uncomfortable if they face a 3D virtual audience within a CAVE) (BH_VRTechE). Furthermore, if a trainee is used to good computer graphic and the virtual audience is rendered with low speed and in poor graphic quality, this could impact on how seriously they take the application (BH_VRTechE).

Another limitation regards the viability of the simulation outside a laboratory. A few technology experts had experience with systems that control virtual humans and explained that they are very time consuming and usually too expensive to be created outside the laboratory. Moreover, they are usually made from leftover software and hardware components of other projects. This again is a time consuming procedure, because they have to be matched to work properly together (RK_VRTechE).

With regard to interactivity, it's very difficult to make virtual humans speak (DT_VRTechE), hence verbal feedback could be problematic if it's not given by the trainer as an avatar with a microphone. Moreover, a virtual human is still "a little dummy", to quote one of the experts, because "it's difficult to make the computer understand a real human" (DT_VRTechE). Also, misplacing the audience in a unfitting environment or in a simulated environment of a lower graphic quality than the virtual humans would also bring discrepancies that could affect user performance (BH_VRTechE).

Another issue is to imagine that the virtual reality simulation could replace a real training session or interaction with a trainer. Although self-help online programs for speech anxiety treatment gave good results (Botella et al., 2010), it is yet unclear whether they would work with public speaking skill acquisition. Further risks stem from intrusive sensors for physiological measurements which could disturb users during performance (TK_VRTechE). One public speaking expert also mentioned her concerns with predictability of simulation–if trainees get used to the simulation and start understanding how that is controlled by the trainer they might get used to it (habituate) and might not respond to it properly after a while (RM_TE).

CHAPTER 3.5

Conclusion

The research study reported here represents the first step in addressing the complexity of a virtual audience design concept. It followed the audience research model which, inspired by the communication studies view (Verderber et al., 2008) gives audience the central role. In the view of communication studies, audiences are a central element in the speech preparation and delivery phases. Both speakers and listeners participate in the speech planning and speech making process and speakers have to know their audiences and adapt quickly to their expectations if they wish their speech to be effective (Verderber et al., 2008).

The current study approached various audience characteristics as experienced and envisioned by trainers, reviewed interactive capabilities of listeners, the group dynamic features of a public, as well as focused on the training procedure. With regard to audience characteristics, the highlight was on audience personalization as method of endowing virtual humans with familiar faces that speakers would recognize. The next research topic was interaction, which was researched from three different angles, interaction as feedback session at the end of a speech, speaker-listener interaction during a speech, and human-computer interaction between a speaker and a virtual audience. The third topic addressed behaviors that appear at group level and spread across the crowd.

The empirical studies were conducted in an exploratory manner, using expert interviews of two complementary groups–public speaking trainers and VR experts. These highlighted what is important from a training perspective and how it can be technically achieved. The data was analyzed according to the coding procedure proposed by Gläser and Laudel (2013), both deductively and inductively. The initial codebook contained three main categories with a few subcategories. At the end of the data analysis, the codebook contained in total 605 categories,

subsumed into seven main categories, further divided into multilayered subcategories.

A field observation of a five-hour speech practice session in a public speaking training seminar was also conducted. The goal was to shed light on the less discussed procedural aspect of training (e.g., at what point does a trainee actually practice the speech, how does this practice look like, and how does it accommodate feedback sessions). Data from the field observation was summed up and presented together with the findings from the interviews.

Table 3.19 represents the original frame of research of this study with the additions obtained through the expert interviews. Following, main results are briefly summarized.

Table 3.19
Frame of research

Topic of research	Components	Research findings
Audience customization (Hilmert et al., 2002; Jackson & Latane, 1981; Maclntyre & Thivierge, 1995)	Audience familiarity;	Known or unknown audience members; Other findings through inductive analysis: Virtual audience characteristics: age; sex; occupation; status; place of origin; outfit; religious affiliation, political views, hobbies. Virtual room suggestions
Interaction (Levasseur et al., 2004; Lucas, 2012; Poyatos, 1985; Watzlawick et al., 1967)	Speaker-listener interaction; Human-computer interaction; Interaction as feedback session.	Verbal and nonverbal audience responses; sensor examples; during and after speech sessions. Other findings through inductive analysis: Graphic user interface with: virtual audience setup and audience control options, choice over modality (verbal and nonverbal), choice over feedback type (content-related or style-related)
Group dynamic (Forsyth, 2009; Ulicny & Thalmann, 2001, 2002)	Contagion; Variety.	Disturbing factors; spreading inattentiveness; randomization rules
Other findings emerging through inductive analysis		Training procedure: speech practice planning, speech and feedback session duration Limitations and risks of virtual audience design: finances, usability, design quality.

3.5.1 Summary of findings

3.5.1.1 Audience personalization

With regard to personalizing audiences with familiar faces, trainers usually go with the preference of speakers. These are free to chose any imaginary audience, even if in reality they speak to their fellow trainees. This rule doesn't apply to speakers who prepare for public speaking competitions and who are particularly encouraged to train in front of unknown people. For them, a virtual audience that looks totally different than the people they know is most helpful. Technically speaking it's also possible to endow a virtual audience with the facial features of people from real life. However, it's very difficult to animate them properly, without avoiding the uncanny valley.

Audience personalization belongs to the first level of the virtual audience research model. Familiarity was conceptualized as similarity in physical characteristics, therefore, the virtual audience ought to include at this level audience characteristics of real persons. This could require a close collaboration between the trainer who does the setup of the audience characteristics and the trainee who provides the necessary details about the real persons they wish to simulate. Familiarity in this sense affects as well the technical attribute "personification", because familiar virtual humans would also have to embody the personalities, moods, and idiosyncratic movements of their real counterparts.

3.5.1.2 Interaction

There is extensive interaction taking place during training in the feedback sessions at the end of the speech practice. These sessions are usually long and thorough. Experts are interested in having virtual audiences which can provide feedback on speech content. The feedback tone is usually positive and supportive of the speaker in training sessions, a feature that can also be implemented in the virtual character's feedback.

During the speech, interaction means audience reactions towards speaker and a speaker's response in return. Trainees are taught to master two important techniques: acknowledging what the audience is doing and reacting properly on the spot. A virtual audience would be particularly useful as training tool, because it could be used to illustrate behaviors, teach trainees to acknowledge them (is someone from the

audience doing something that should worry me, as speaker?), and train immediate responses (e.g., having a joke prepared, in case audiences get bored, or even take a break from the speech if people are tired).

The prospect of interaction between a speaker and a virtual audience yielded two major directions: towards audience autonomy and towards audience control. Experts would prefer easily controllable audiences, in real time, through buttons. However, there are also many sensors that can help implement autonomous audiences. Autonomous audiences that can further understand speech content could provide extra structured feedback on how well speakers performed. One could therefore devise virtual audiences with various complexity levels, from simple ones that offer feedback limited to a few movements, such as eye gaze, to highly complex ones which could ask content-related questions.

Interaction constitutes the second level of the audience research model. Initially concerned only with speech delivery and audience's responses, interaction expanded to encompass the feedback session at the end of the speech. Although structurally similar in terms of verbal and nonverbal communication, feedback sessions are described by training experts as heavily focused on verbal input. This aspect may change the technical requirements and expectations of virtual humans (technical attributes) by demanding high artificial intelligence capabilities to sustain understanding and production of natural language. On the other hand, if technology doesn't allow virtual humans to make reasonable comments on a speech, these could be used then only during a speech where they don't interfere verbally.

3.5.1.3   Group dynamic

Trainers described the kind of group dynamic they've encountered and there are three types of triggers they identified: the speaker, someone in the audience, and someone or something outside the speech setting (e.g., on the street). When one person becomes restless, it is very easy to disturb more and more people and trainees are also taught several audience management skills to deal with this kind of behaviors. Group dynamic is further researched in the second study of the dissertation, where a real student audience was observed and behaviors that spread could also be noticed.

Group dynamic belongs to the first level of the research model closely connected to the VH autonomy attributes. The behaviors listeners

could manifest without the influence of the speaker, based only on what they perceive from other listeners or inputs in the virtual environment (e.g., behavior contagion) translate in the virtual environment into virtual characters that are responsive to their simulated world.

3.5.1.4 Training procedure

The training procedure represents a vast category which rests on the interviews and on the speech practice observation. It contains organizational data, such as speech durations (can range from 1-8 minutes), speech types (formal or informal), and trainees' demographic profiles.

Understanding the training procedure is relevant for identifying the potential of the virtual audience for communication training altogether. Even if in general free to chose their audiences, speakers establish together with trainers their speech goals and the audience they would have to address in real life and this could represent the first phase in virtual audience customization, both in terms of looks and of behaviors. The speech observation seminar revealed that speech practice takes place only after a thorough theoretical introduction. At that point, the virtual audience could be used to illustrate relevant audience and behavior typologies. Speech practice is usually very short, but feedback sessions are extensive (up to one hour). Therefore, the virtual audience could cover both these sections. However, if virtual audiences are meant to be used in the feedback sessions, they ought to have extensive cognitive abilities, to participate in the Q&A sessions and to give sensible feedback to the speaker both in terms of content and presentation style.

The findings on the training procedure reflect existing literature that mentioned speech preparation, delivery, and feedback phases (Beebe, 2007; Hargie, 2006). Moreover, the collaborative work between speaker and trainer is similar to practices found in phobia treatment, where CBT procedures require that patients elaborate together with therapists lists of feared situations (e.g., Wallach et al. (2009)). However, in the speech training situation, it appears that trainers can take a more active role in what situations should be trained (by asking participants to take up role playing of random speech situations), compared with phobia treatment, where patients are the main drivers of played out fear hierarchies (Deacon & Abramowitz, 2006).

Training procedure belongs to the third level of the audience concept, as it can be seen as part of the larger training program. Although the findings can't be generalized due to the single case analysis, one could safely state that a standardized speech duration could be used across various populations. A study by (Ayres, 1996) proved that the time rehearsing with an audience didn't differ significantly between low and high apprehensive speakers. However, high and low apprehensive speakers differed in the amount of time they spent preparing notes, analyzing the audience, and silent rehearsal (Ayres, 1996). These differences suggest that the actual duration of a scenario with the virtual audience could be the same for users across the whole spectrum of anxiety (low to high; state and trait), whereas preparation phases and audience analysis might require customized attention from the trainer. The authors suggest that high anxious speakers spend proportionally more time preparing notes than practicing a speech and analyzing the audience and this imbalance could be corrected with proper training. For instance, the virtual audience could be implemented to showcase specific behaviors and help avoiding speakers invest more time in communication-related activities than they would usually do.

3.5.1.5 Virtual audience characteristics

Several audience demographic features were mentioned, such as age, sex, race, religion, profession, status, and outfit. Apart from customizing these characteristics, experts mentioned various audience sizes and virtual locations they could populate.

Regarding audience characteristics, these were divided into verbal/nonverbal and positive/negative/neither nor. Two matrices occurred, one with audience behaviors independently of what a speaker does and one with reactions to the speaker's input. Audience behaviors appeared in many sections and they represent a major point in training. For this reason, they were further investigated in the second research of the present dissertation.

Another element regarding groups was variety. Realistic audiences are expected to have individuals look and behave in a diverse fashion. To achieve this, one can randomize behaviors across the audience, their time of occurrence, intensity, order, and their idiosyncratic features specific to each listener. Randomization rules can be used by trainers to organize the characteristics and behaviors of a virtual audience before the speech begins, in the setup phase (e.g., behavior type, time of

occurrence, intensity, etc.).   The purpose is to avoid redundancy of
virtual human characteristics, which could threaten simulation realism.

These audience characteristics make up the majority of features under
the first level of the research model.   They address mainly physical
attributes that fall under demographic characteristics. The first level
included also cultural background and psychological characteristics
relevant for the effectiveness of a speech (Verderber et al., 2008).
Although not inquired here, if needed, personalities and emotions can
still be depicted in a realistic manner through the technical attribute of
"personification" (first level) with the help of existing emotion models
(e.g., the computational model EMA used by Gratch and Marsella
(2004)).

3.5.1.6   Virtual audience setup

Among the discussed technical features were methods to customize and
control the virtual audience, such as having a possibility to chose how
many men and women are in a group, what's their hobbies, what they
like and don't like, etc. This should occur in a simple manner, just
like in the case of interaction control discussed above, through simple
buttons or other graphic user interface options, such as drop-down
lists. Moreover, an audience should be able to be customized before the
speech starts and also be controlled once the speaker starts practicing.
Trainers would like to be able to trigger various audience reactions, and
these should be already available in the system and easily set during
the speech.

The virtual audience setup draws back to both the first and the
second level of the audience concept. Setup is a proposed procedure
through which the trainer selects among the human-like attributes of
the virtual characters (first level) those ones that are most relevant
for the trainee and the training purpose. In the training procedure,
this can be seen as the first step before the actual speech practice
session begins. Therefore, this becomes a necessary condition for the
start of the speaker-audience interaction (second level). Grillon et
al. (2006) mentioned this procedure in their study on social anxiety
treatment.   Another element that can be set up before the speech
begins is the virtual location where the speech is going to take place.
A particular case for virtual room setup is given by Rizzo et al.
(2002) who suggest using a virtual room to assess memory processes
by placing various objects that users have to notice and recall. Other

studies simulate consultation rooms in communication skills training for medical students, and they adopt usually a standard consultation room setup, sometimes as extension of the real consultation room (Johnsen et al., 2005). Whereas room setup in medical communication skills does not affect the displayed characteristics of the virtual patient who is usually completely visible and seated (Johnsen et al., 2005; Kenny et al., 2007), the situation is different for public speaking simulations. Here listeners could be only partially visible due to their seating in the room (one behind the other). Therefore, the virtual room setup could influence the displayed characteristics of the virtual listeners, independently of the initial choices selected by the trainers.

3.5.1.7   Limitations and risks of virtual audience design

Limitations and risks of design were pooled together and presented in a separate category to make it easy to pinpoint problematic aspects of design. Several experts commented on the risks of designing such a virtual audience for training purposes. Although supported by all experts, a training solution with the interactive capabilities expected by speech coaches is hard to acquire with the actual technology, especially when one considers the financial aspects. Other risks consist of using insufficient computing power to properly animate and display virtual characters (first level of the research model), intrusive sensors for achieving interactivity which could disturb trainees during their speech (second level of the research model), as well as labor intensive design that can interfere with the training practice program (third level of the research model). Limitations and risks appear pervasive across the design levels and future research should address them separately along each design step.

3.5.2   Discussion

The study revealed the way trainers see a virtual audience and how they would like to make use of one. All VR technology experts had previous experience with virtual reality and even audience simulation. By comparison, trainers who have never worked with any VR application were much more optimistic with regard to what they would like to have. Their creative thoughts regarding what audiences should do, such as give standardized feedback on how well someone performed, brings a fresh view in audience design. So far, virtual audiences have

extensively been used in phobia treatment. This oriented the design efforts towards creating audiences that provoke anxiety responses and are not constructed to deal specifically with speech improvement.

Technology experts enumerated several design risks that ought to be considered, such as high design costs, experience of users with high fidelity image, and the high expectations they would have from the virtual reality training. Any solution should be considered with care and the design process ought to be kept within realistic requirements and expectations.

Public speaking training is a procedure that can apply to anyone, independently of their speech anxiety. Therefore, audience customization options go beyond fear eliciting factors, and orient themselves after elements that are relevant for speakers in their endeavor to improve speech performance. The question is not anymore how not to be afraid to speak in front of a person, but how to address that person in a proper way and achieve the communication goal. Knowing and understanding audience variety is important now primarily because this is how a speech can be customized to address the needs and expectations of listeners, and not so much because the speaker has to get acquainted to and feel less anxious in front of diverse groups.

Experts enumerated many behaviors that they've encountered in audiences and these have been explained here with the help of a matrix, as positive/negative or verbal/nonverbal. A careful consideration of these behaviors revealed that they are extremely heterogeneous (e.g., yawning, murmuring, changing posture, taking notes). They contain various nonverbal and verbal modalities (e.g., clapping, booing, asking questions, using interjections).   In order to be able to categorize them into customizable options, it's important to structure them along clear-cut lines and understand what differentiates them.   For this reasons, audience behaviors became the subject of the next study in the dissertation. Another type of behavior that was further inquired in the next study was group dynamic. In the interviews, experts only mentioned what behaviors they've noticed spreading across the audiences. The information was useful and yielded a few ideas for future research, such as identifying the triggering factors. Therefore, the analysis of a real audience appears as a more fitting method of understanding how such behaviors occur.

Another important aspect was degree of audience control. Trainers would like to be able to tune the audiences in real time to match a speaker's performance. This control calls for a robust system that can support complex changes within audience content. Moreover, it also raises the question on work load: how difficult is it for trainers and is it even possible to monitor the speaker and actuate audience responses at the same time?

In order to address as many performance levels as possible, a solution could be to implement audience packages. There could be a beginner's virtual audience that has a standard package of customizable features, such as audience size, sex ratio, and positive/negative nonverbal behaviors. As a speaker's needs expand, the virtual audience could become more and more complex and include new packages of customizable features, such as backstories for each character and the option of verbal feedback. These options ought to be considered with care, as there are many ways to customize virtual humans, and the most relevant elements should also be discussed in future studies with both trainers and trainees.

Little direct reference was made towards elements specific to immersive environments, such as immersion, interaction fidelity, and presence. Experts were informed from the beginning that the questions address virtual audiences implemented in 3D environments. The fact that these topics were not particularly discussed with experts was a conscious decision of the author. The purpose of the audience design concept was to offer a preliminary list of features to be implemented in the virtual audience. Immersion within a virtual audience scenario, the feeling of presence, and interaction fidelity are elements that one should always consider from the beginning in the design process of VR applications. However, they are ultimately related to how users (speakers) experience a system and a simulation. Therefore, a sensible way to address them is through user studies where the virtual audience concept can be tested in practical speech applications. For now, it can be hypothesized though that the level of audience customization that emerged from these findings is in line with the model of presence in learning environments proposed by F. Mantovani and Castelnuovo (2003). The tool suggested here covers some of the most important factors that support presence: it can be adapted to the needs of trainees, it suggest highly interactive content which the trainee can control (via verbal and nonverbal feedback), and it recreates the actual

training situation experienced in real life for optimum skill transfer. Prototyping such a virtual audience could further support research on social presence in learning environments with artificial intelligence. Furthermore, it could help advance the state of the art of virtual learning environments in the communication field, by following the CID and CIX paradigms in a broader spectrum of interest across various disciplines.

All relevant features of the virtual audience that occurred through the expert interviews were also complemented by the detailed behavior analysis in the second study and ultimately integrated into the virtual audience design concept.

3.5.3   Limitations and outlook

The study synthesized a large pool of information gathered from seven expert interviews and one field observation. One of the major limitations is of methodological nature and refers to the small number of training experts compared to virtual reality ones (two and five), as well as the altogether small sample size. Trainers showed great enthusiasm at the idea of a virtual audience they could use in their seminars. Moreover, they also provided ample descriptions of how they would like the virtual audience to be. However, in future studies, trainees should also be included as one of the major stakeholder groups. They could express their own needs and expectations from a virtual audience as well as assess its precise role for their own skill development.

Originally, the study was meant to address three types of experts: public speaking trainers, designers of virtual audience applications, and programmers in the VR field. However, in the interview analysis phase it was difficult to crystallize separate answers from all these three perspectives. Therefore, the answers from designers and programmers were analyzed together. These two groups formed the VR technology experts. This can also explain the discrepancy between the interview guidelines, which were meant to approach the same topics from different perspectives, but were ultimately analyzed in a complementary manner and not through a comparative one. Moreover, future studies could use a different methodology, with more focused questions. The semi-structured approach had the advantage of letting experts reveal topics they found interesting, but it made it also difficult to have

the researched topic covered equally. Another limitation regarding the interviews is that they were conducted simultaneously. A better approach would have been to conduct the interviews with trainers first, to have a better impression on what is actually expected from a virtual audience and then discuss those precise aspects with VR technology experts.

The interviewed training experts were specialized in public speaking, and their needs and expectations revolved around teaching trainees of the diversity of potential audiences. However, the study here showed what is technically doable and what are the potential gaps in research with regard to having highly customizable virtual humans. Many actual studies are dedicated to virtual human autonomy, whereas agent customization or systems that deal with customized agents are only emerging (e.g., virtual patients (Kenny, Parsons, & Rizzo, 2009)). This opens up the opportunity to turn to user interfaces and systems that can help customize virtual humans, as well as do this outside the laboratory, on a commercial platform.

The discussed data was ample with regard to training procedures and expectations but less detailed on specific technical aspects. This can be partially explained by the fact that the questions on technical aspects were of general nature. The intention was not to get into too many technical details, but to answer rather general questions on the main researched topics and also be able to bridge the answers across disciplines. Future studies could use a different research design, where experts from relevant areas participate in focus groups or task analyses and can discuss in more detail how to fulfill the requirements of training sessions with existing technologies.

4

Audience behaviors research study

CHAPTER 4.1

Introduction to audience behaviors

Virtual audiences and singular virtual characters represent an important tool used in speech anxiety treatment or in communication training applications that address one-to-one communication (Kenny et al., 2007; Pataki et al., 2012). Previous virtual audiences used for anxiety treatment have been designed to display several nonverbal behaviors, from relaxed and friendly, to bored and negative ones (Pertaub et al., 2001; Slater et al., 1999; Wallach et al., 2011). In case of anxiety treatment, the choice of behaviors has always been congruent with the treatment purpose–that of eliciting different levels of anxiety with the help of various levels of displayed friendliness, unfriendliness, positiveness, or negativeness. One can regard the design and display of such behaviors as rather normative. Because of their restricted treatment purpose, they contain clear-cut behaviors that make audiences seem homogeneous in their responses. This poses a bias risk towards behaviors that are displayed at their intensity peak (fully blown) and with little variation among individuals.

In real life, behaviors are not always obvious or easy to observe and they sometimes appear with different intensity (Hess, Blairy, & Kleck, 1997). Moreover, as training experts stated, groups can be unpredictable and the change in their behavior as well as its heterogeneity ought to be properly displayed. This variety of behaviors is important from the perspective of public speaking training procedure, because trainees are required to adapt to and perform in accordance with various audience moods, cultures, ethnic and religious backgrounds, political views, speech settings, etc.

This present study addresses the first level of the audience research model and focuses only on listener's manifest behaviors during a lecture. The inquiry was based on what experts discussed with regard to the diverse audience behaviors they have come across in their work. In accordance with their expectations, a virtual audience

should accommodate various physical characteristics and accompanying behaviors. Having virtual listeners display a few categorical behaviors has proven successful for speech anxiety treatment, where subjects eventually managed to lower their anxiety levels through exposure to a virtual speech situation (Pertaub et al., 2002; Wallach et al., 2009). However, for people who present and pay attention to their listeners, who try to convince them of their arguments, and engage the audience in the speech, behavior diversity and flexibility might prove to be an asset.

To reach this goal, the study focused on the manifest nonverbal behaviors of one relevant audience behavior from public speaking perspective–that of audience interest in the speech or attention towards a speaker (Turk, 2004).   Apart from attentiveness, its opposite–inattentiveness was also considered, to give a balanced view of what audiences do while the speaker is performing. The first question was RQ1: *How does attentiveness and inattentiveness manifest nonverbally within an audience?* In spite of focusing only on two major audience reactions, the exploratory nature of the research is meant to identify the underlying diversity of nonverbal behaviors that make audiences look like they are paying attention or not. The second research question was RQ2: *What nonverbal behaviors in the audience trigger nonverbal responses from fellow audience members at group level and how do such behaviors spread?* Observable audience responses prove that audience members can influence each other nonverbally (Hylton, 1971). The question posed here wishes to complement the findings in the first study with behaviors observed in a natural setting.

The study has a bottom-up approach on audience design because it looks at how behaviors occur in a real life presentation situation. A student audience was observed during a lecture at Technische Universität Ilmenau and the observation focused on the nonverbal markers of attentiveness and inattentiveness.   Students were not required to respond to any questions on how much they actually payed attention to the lecture, therefore no inferences can be made on why they behaved the way they did. The proposed observational procedure helps the design process be less normative:

- it ensures that designed behaviors in the VR setting are congruent with those occurring in real life; peaks and lows of frequency in

animated behaviors, as well as behavior dullness can be this way empirically grounded;

- it ensures that behaviors are observed in a multimodal fashion: head, torso, and hands are coded simultaneously and form one behavior unit. This unit can be transferred to a virtual character that performs the same movements with the exact same number of nonverbal modalities, in a similar fashion like a real person;

- nonverbal behavior duration resembles the experienced in real life.

The next chapter comprises an introduction into the theoretical background of nonverbal behaviors that can be subscribed to attentiveness and inattentiveness. Next, the methodology of the study is introduced and presents the category system, the development of the code book and of the coding interface, followed by the affiliated coding rules. Results are presented afterwards, with brief summaries after each major behavior discussed. The chapter ends with a discussion of findings and with limitations and suggestions for future studies on real and virtual audiences.

CHAPTER 4.2

Theoretical background on nonverbal audience behaviors

4.2.1 Audience definition

Audiences are a type of collective which deliberately gathers to observe a happening at a certain point in time, such as a group of people gathered at an exhibition or at a performance (Forsyth, 2009). They belong to the same taxonomy as street crowds, queues or mobs and differ from other forms of groups, such as families, teams, or work crews in that they are less structured and lack well defined and stable membership (Forsyth, 2009; Rapee & Heimberg, 1997).

In the domain of social anxiety, phobics view audiences not only as a group of intentional observers, but as evaluative entities who have expectations over their performance and whose evaluations impact the performance quality directly (Rapee & Heimberg, 1997). Speech phobics for instance recall less about the environment they perform in, such as the characteristics of the audiences and they focus more on their own behaviors to meet the expectations they believe the audience would have from them, compared to low anxiety speakers (Daly et al., 1989). The fact that low anxiety speakers devote more attention towards the audience and environment than high anxiety ones is a first hint towards training solutions with audiences that act in a realistic manner for speakers to acknowledge their presence and actions.

Speaker and listener roles are actually hardly clear cut and anyone who performs the act of communication is simultaneously a speaker by performing and a listener by attending and evaluating the communicated information (Bakthin, 1986). In a simplified manner, a speaker is the person who has the floor, whereas the listener is the person who waits for their turn (Heylen, 2009). In public speaking training seminars, mostly everyone gets to be speaker (RM_TE, CS_TE). Group diver-

sity may vary highly in a training class, and experts mentioned speakers who ranged in age from primary school to elderly, from management employees to management aspirants, from people who simply wish to improve presenting skills to people who wish to win public speaking competitions. When not performing during the training sessions, participants fulfilled the other role, of audiences.

4.2.2 Audience behaviors

Audience behaviors depend on many factors, such as external factors determined by the location of speech events (Forsyth, 2009) and event conventions (Blumer, as cited in Forsyth (2009)). People in gatherings such as audiences perform all kinds of behaviors, such as moving (e.g., searching for their assigned seat), verbalizing through language forms, or vocalizing through paralinguistic forms (e.g., cheering) (Forsyth, 2009; Knapp & Hall, 2010; McPhail, Powers, & Tucker, 1992). When they attend organized events, audiences are bound to event conventions, such as accessing the hall based on seating order and on social conventions (Blumer, as cited in Forsyth (2009)). Seated audiences in particular can be bound to their location and their behaviors can be limited to a smaller repertoire, which nonetheless can still contain various torso movements and postures, as well as arms and legs movements. McPhail (2014) focused on face, voice, hands, and legs for classifying the studied movements of people in groups. Other researchers add further components of human anatomy, such as torso and head, as well as complex movements that comprise simultaneous activation of separate anatomic parts (e.g., playing on a mobile device (Mann & Robinson, 2009)).

Other factors that may determine how audiences behave belong to the human context that can again be divided into (1) factors that are audience-bound (e.g., audience cultural background, psychological setting, familiarity with speaker and speech topic (Verderber et al., 2008), as well as the behaviors of other audience members (Hylton, 1971)) and (2) factors that are speaker-bound (e.g., speech quality and performance (Beebe & Beebe, 2010; Verderber et al., 2008)). Some of the interviewed experts also highlighted the role of demographic profile in the way audiences behave: younger audiences tend to be very active compared to older ones, people from conservative cultures (e.g., Russia,

Nigeria) tend to have very little emotional facial expressions (CS_TE, RM_TE, BH_VRTechE).

Another way to look at audience behavior is through the valence of their behaviors. Depending on their level of interest in the performer, audiences can be divided into three types: (1) supportive, (2) neutral, and (3) negative (Butler & Baumeister, 1998). Supportive audiences have favorable attitudes and encourage performers, through identity link or social bond (Cialdini et al., 1976). Neutral ones have little stake in performers' outcome and refrain from specific manifestations, whereas negative audiences are unsupportive and discouraging towards performers (Cialdini et al., 1976). This categorization has to be considered with caution. It may be easy for the performer or for external observers to categorize audiences based on what those audiences actively do, or, as in the case of speech phobics, based on what performers *think* audiences do (Daly et al., 1989). For an audience member though, this might not be the case at all, as one trainer mentioned in the interviews (RM), especially when people are determined by circumstances to attend a speech and act bored not because of the speaker, but because of just having to be there and listen. In order to avoid any type of confusion between what audiences display and what they actually feel, only manifest (Compton, Love, & Sell, 2012) or displayed behaviors will be considered here. As mentioned in the introduction of this chapter, the reasons behind displayed nonverbal behaviors have not been inquired.

In one-to-one interaction, affect vocalizations such as "mhm, mhm" are a common backchannel and can have positive meanings when uttered in the proper pitch or intonation (Scherer, 1994). There are few such vocalizations during public speeches, and, depending on the speech format, speakers may be met with applause, booing, whistles or cheers (Clayman, 1993). Other audiences can be more active and even cruel, to quote stories presented by a public speaking expert (RM) who spoke of people leaving the room earlier, coming in late or laughing so hard at the performers that they would burst into tears.

As seen in the expert interview study, public speaking experts (RM and CS) also stated that speakers have to learn to be critical of audience reactions and always consider whether their reactions are actually triggered by the speech or by factors external to the speech situation. Trainees should adapt to various audience attitudes that can span from

total neutrality (which can be very unsettling to experience, according to RM_TE), to very inattentive ones, especially when they play out group dynamic and engulf the whole present audience (CS_TE).

Two types of audience behaviors have been selected in previous studies on fear of public speaking: positive and negative audiences. For several virtual audiences, positive or negative displayed behaviors have been selected from literature and implemented in the virtual human design (Pertaub et al., 2002) or the public speaking setting has been created together with the trainee or patient based on their own hierarchy of feared situations (Grillon et al., 2006). Other studies recreated virtual audiences based on observations of real audience members and used randomization algorithms to recreate credible and realistic virtual audiences (Kang et al., 2013). The present study on real audiences starts as well from an observation, but is focused only on attentive and inattentive behaviors. The selection of only two behaviors was advantageous because it allowed an in-depth analysis of nonverbal manifestations.

4.2.3   Attentive audiences

This category aggregates a few nonverbal behaviors that depict an interested audience in the speech as well as approach behaviors towards the speaker. The subchapter introduces the nonverbal manifestations of attentiveness and immediacy, since both have a positive valence. Immediacy has been researched as a desirable behavior for teachers that fosters student attentiveness (Richmond, 2003). Since they both share similar nonverbal markers, such as eye gaze and body orientation towards the source (Hale & Burgoon, 1984; Myers & Ferry, 2001; Norton & Pettegrew, 1979) they were both considered in the present study.

4.2.3.1   Attentiveness

Attention is one of the most studied concepts in education and public speaking and can be defined as "a state of focused awareness on a subset of the available perceptual information (Gerrig & Zimbardo, 2002). This goal-driven attention is also called endogenous attention (MacLean et al., 2009). Whereas attention is concerned with reception and cognition, attentiveness was defined as a mixture of verbal and

nonverbal signs people use to indicate that they are noticing and understanding someone's message (Norton & Pettegrew, 1979).

People can select among sensory inputs (Chun & Wolfe, 2001) and orient their attention towards the direction of preferred stimuli (Frischen, Bayliss, & Tipper, 2007).   An overt orientation (Posner, as cited in Frischen et al. (2007)) means that a person orients their sensory receptors such as gaze and/or head towards the stimulus, to allow better perception. In a presentation or public speaking situation, audience members are usually seated and face the speakers. Presenters can use these nonverbal cues to evaluate how much attention do audiences pay.

There are many markers of perceived displayed attention in interpersonal communication, such as body orientation, smiling, nodding, leaning towards a speaker, direct eye gaze (Norton & Pettegrew, 1979), and mimicry (Gueguen, Jacob, & Martin, 2009). For groups of listeners, attention has been related to direct gaze towards a speaker, little body movements, and note taking in study contexts (Wilson & Korn, 2007).

Virtual audiences display attention through body orientation towards speaker, gazing patterns that follow the speaker, as well as smiles and nodding (Pertaub et al., 2002). Figure 4.1 depicts a snapshot of the virtual audience developed by Kang et al. (2013) indicating various attentive postures.

The observational study wished to identify what other nonverbal behaviors can be assigned to attentiveness based on gaze and body

Figure 4.1. Attentive virtual audience (Kang et al., 2013, p. 335)

orientation towards speaker that can ultimately be implemented in a virtual audience.

4.2.3.2 Immediacy

Another concept often researched with student groups is immediacy (e.g., Christophel, 1990; Richmond, 2003). Immediacy is represented by verbal and nonverbal behaviors that increase or decrease closeness between people and signal approach-avoidance tendencies (Hale & Burgoon, 1984). According to Mehrabian (1971) who introduced the term, 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" (Baringer & McCroskey, 2000, p. 178). In the education context, where teachers, lecturers, or professors take up the role of a speaker and the pupils (students) that of the audience, teacher immediacy facilitates information retention and motivates students (Christophel, 1990). Immediacy manifests nonverbally through increased eye gaze, positive facial expressions, physical proximity, and body lean towards the speaker (Hale & Burgoon, 1984; Myers & Ferry, 2001). Even though immediacy was studied more for the speaker and less for the audience (Baringer & McCroskey, 2000), similar nonverbal behaviors have been already observed to occur within groups as well (Kang et al., 2013): during lectures, students have eye contact with the presenter and lean forward, as if to reduce the perceived distance between them.

Both literature on attentiveness and immediacy provide useful nonverbal markers for positive attitudes in interpersonal communication. Immediacy is relevant for the present observation study because it complements the nonverbal markers of attention mentioned above. Studies on groups yielded similar behaviors that engage the same modality categories: eye gaze, posture orientation, or head movements. For the current research, the following nonverbal markers were chosen for the observation: eye gaze behavior, nodding, front-oriented body posture (Richmond, 2003), and note taking (Wilson & Korn, 2007).

Several behaviors such as smiles or frowns were removed from the original list due to the distance between camera and the recorded person, which impeded a clear observation of facial expressions. For simplicity reasons, only the term attentiveness has been used further and it denotes behaviors that belong to both attentiveness and immediacy.

4.2.4 Inattentive audiences

This category introduces the nonverbal manifestations of two behaviors that have negative connotations: involuntary attention and boredom (also called here deliberate inattentiveness). Their negative connotation stems from the fact that audience members that display them adopt an avoidant behavior: they look somewhere else and are busy with activities unrelated to the speaker (Mann & Robinson, 2009).

4.2.4.1 Involuntary attention

Involuntary attention represents an exogenous or stimulus-driven attention away from an original gaze orientation towards a stimulus, determined by the occurrence of events, such as abrupt movement in the visual periphery (MacLean et al., 2009; Remington, Johnston, & Yantis, 1992). This occurs when a person doesn't allocate resources on purpose to take notice of an event happening, but merely responds to its stimuli, in a bottom-up approach, for instance when something appears to move in the corner of the eye and the person turns their head to look in that direction (Eimer, Nattkemper, Schröger, & Prinz, 1996). For the present audience observation, involuntary attention was conceptualized as the allocation of attention resources to a different stimulus than the speaker. This translated into speaker-avoidant behaviors such as averted gaze away from the presenter and averted body postures away from the frontal orientation towards the speaker. Such shifts in attention become involuntary attention when they are triggered by events happening around the coded listener, such as another student who distracts the coded person or any other event that enters the auditory and visual field of the coded person.

The practical value of including involuntary attention in the analysis was to justify the influence that external events can have on a person's attentiveness towards the speaker. Such events were mentioned also by experts in the previous studies and can refer to a person entering the room mid-lecture, a person initiating a conversation, a person starting to laugh loudly, or people showing or passing on objects. Other kinds of speaker-avoidant behaviors that denote reallocation of attention away from the speaker are presented below in the broad category of boredom. From a technical perspective, people who distract other audience members can be regarded as triggers for involuntary attention and can serve as models for virtual humans that fulfill the trigger role

within the virtual audience. In virtual reality applications, involuntary attention could be equated with having conversing virtual humans: a virtual character engages another one into a conversation and distracts the former from paying attention to the speaker.

4.2.4.2 Boredom

Everybody has experienced it, yet there is little consensus in literature on the definition of boredom. A thorough literature review helped researchers agree that it involves negative emotions (e.g., lacking interest and enjoyment in Jagacinski and Duda, as cited in Vogel-Walcutt, Fiorella, Carper, and Schatz (2011)) and low arousal (e.g., acute under-stimulation in Mastro et al., as cited in Vogel-Walcutt et al. (2011)), either separately or together: "State 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" Vogel-Walcutt et al. (2011, p. 102).

People employ different coping strategies to avoid the unpleasant state of boredom (Csikszentmihalyi, 1975). Students particularly make use of electronic devices or start talking to each other during class: they play games on the mobile devices, send text messages, doodle, switch off, daydream, talk to fellow students next to them, even leave mid class or at break time (Mann & Robinson, 2009). Object manipulation or object adaptors usually involve rubbing, playing, or actuating inanimate objects. When employing such behaviors, people shift their attention away from the speaker and appear to be inattentive. Therefore such behaviors have been used in the codebook to mark deliberate inattentiveness.

Looking back at previously designed virtual audiences, in the negative setup, virtual humans were animated to display highly conspicuous negative behaviors, such as putting their legs on the desk or fall asleep (Pertaub et al., 2002). Less conspicuous negative behaviors inspired by real audience observations (critical and bored scenarios) showed virtual humans with crossed arms and looking down to desk, as seen in Figure 4.2 (Kang et al., 2013).

Figure 4.2. Bored virtual audience (Kang et al., 2013, p. 335)

In order to enrich the behavior catalog for virtual humans, the observational study has considered all behaviors that contain averted eye gaze behavior from the direction of the speaker and of the presentation slides. These behaviors contain not only eye gaze direction, but also torso, arm and hand postures, as well as specific activities that people perform while they are not looking towards the speaker.

To wrap up, involuntary attention is triggered by a stimulus that grabs the attention of the person towards something different than the speaker, as seen above. The second is performed voluntarily by listeners when they turn away from the speaker and engage in a different activity, such as for instance checking their phone or starting to read something on their laptops, as it happens when people experience boredom. In order to make this differentiation possible in the observation, a triggering actor was coded for each coded behavior. A triggering actor could be: (1) the coded person who decided to perform an action as in deliberate inattentiveness or it could be (2) a person who distracts the coded person and makes them perform an action as in involuntary attention. Moreover, since gaze direction towards a person is a pervasive marker for both attentiveness and immediacy, it was decided to use gaze direction (towards a speaker or in a different direction than the speaker/front of the room) as discriminating factor between attentiveness and inattentiveness.

After their implementation in the virtual audience, such behaviors should pass through the formative evaluation with users (Gabbard, Hix, & Swan II, 1999). Further analyses can reveal if users perceive the

same behaviors as the ones intended (Kang et al., 2013) and by this help calibrate animation intensity.

4.2.5   Neutral nonverbal behaviors

Apart from the nonverbal behaviors associated with attentiveness and inattentiveness, there are other behaviors that have been included in the analysis and which can't be associated with any positive or negative valence unless they are interpreted within a larger context. Such behaviors stem originally from the pretest that was conducted before the actual observational study and represent conversational hand gestures and self-adaptors. Both types belong to the same typology of hand movements, but differ in their degree of lexicalization, with adaptors being least lexical in the sense that they do not sustain any particular meaning in the interaction (Krauss, Chen, & Chawla, 2004). These are self-touching movements that have no connection to the speech, such as playing with the hair or scratching (Beattie & Shovelton, 1999). The function of self-adaptors is an emotional one, because they represent coping mechanisms with own feelings comforting emotional arousal (Ekman & Friesen, as cited in Knapp and Hall (2010)). It is therefore hard to see self-adaptors as markers for attentiveness or inattentiveness in the absence of further cues, such as eye gaze or body orientation relative to the speaker.

Conversational gestures relate to speech through their simultaneity and temporal coordination to speech and share a semantic content to the speech (Ekman & Friesen, as cited in Knapp and Hall (2010)). No semantic content of the observed conversations was analyzed in this observation and conversational gestures were coded only in the presence of observed conversations. They were coded also independently of their typology (iconics, metaphorics, dietics, and beats (McNeill, 1992)) or functions (e.g., suggesting turn taking (Knapp & Hall, 2010)). Usually, gestures are coded between the preparation, stroke, and retraction phases (Efron, as cited in Kipp (2003)). Here gestures were coded between the moment they started until the person put their hand at rest, with no particular attention to their intermediary stages. However, without knowledge of conversation content, gestures were also considered rather neutral body movements that bear no direct association with attentiveness or inattentiveness in the absence of further cues, such as gaze or body orientation relative to the speaker.

4.2.6 Group dynamic

The topic of observable audience responses was approached also in the first study of the dissertation, under the name of group dynamic. In the speaker-audience configuration, there are two interaction types that occur: (1) source-listener interaction and (2) listener-listener interaction (Holtzman, 1970, as cited in Hylton (1971)). This second kind of interaction was labeled *observable audience response* (Hylton, 1971) or "audience-to-audience response" (Kowal, 2009). Hylton (1971) hypothesized that listeners would influence each other in attitudes towards a speaker (manifested through verbal and nonverbal communication) when observing the nonverbal attitudes of people around them. People in an audience were more approving of a speaker if they observed the positive reactions of fellow listeners, such as nodding and smiling. Hocking, Margreiter, and Hylton (1977) show that concert goers can influence each other in their attitude towards a band and stay longer to a concert if other listeners show signs of enjoyment, such as applauding or leave early if other listeners don't pay attention, talking to each other and ignoring the band. No data was provided though about how many people it took to influence the attitude of a whole audience and what percentage that means reported to the size of the whole audience.

Inside an audience, individuals can influence each other in their attitudes and nonverbal behaviors, and these behaviors can spread from one individual to the other across a whole group (Farkas, Helbing, & Vicsek, 2002; Hocking et al., 1977; Hylton, 1971). One example that occurs in large audiences is the La Ola or Mexican waves that occur on stadia (Farkas et al., 2002). They require only a few dozens of people to stand up simultaneously to trigger a mass reaction starting with the people seated next to the initiators (Farkas, Helbig, and Vicsek, 2002).

An article published by Beiseler (2004) mentions knocking as a correspondent of applauding in German universities. Whereas its origin is uncertain, knocking on the desk has been associated with student discipline and has a positive connotation (Harrow & Dziuban, 1974). This behavior occurred at the end of the observed student lecture. Kowal (2009) differentiates between audience reaction types depending on triggering factor. In the case of applause for instance, the ones caused by speakers alone are sudden and burst with high intensity. A gradual onset of applause would indicate an initiator within the audience that is being followed by the others who gradually start

applauding themselves. Booing on the other hand, is usually guided by the observation of what others do, as reported by Clayman (1993).

Another form of contagion is emotion contagion such as smiling, which takes place through direct interaction among individuals and through observation processes of their nonverbal cues (Barsade, 2002). Mimicry (behavior mirroring) is another type that occurs through interaction among people who sit next to each other and who tend to pick up on their partners postures (Knapp & Hall, 2010) or facial expressions (Gueguen et al., 2009). In the present observation emotional contagion was left aside due to the fact that students were scattered across the lecture hall and clustered in small groups of maximum five people in a row. Moreover, the distance from camera to the recorded persons, as well as the camera angle didn't allow an accurate reading of facial expressions for basic emotion identification (Eckman & Friesen as cited in Knapp and Hall (2010)).

4.2.7   Body movements analysis

Attentiveness and inattentiveness behaviors described above appear in a categorical manner, in some cases as given complex movements, in others as simple ones. An example is the case of an object adaptor– having a cup of coffee on the desk and playing with it. The behavior itself is easy to imagine for anyone who had coffee or tea. However, just because this activity is easy to imagine and visualize as a whole, it becomes more complicated to regard it as a group of smaller activities, such as picking up a the cup, rotating it in the air few centimeters above the desk, moving it left and right in the hand, and then putting it down. Figure 4.3 shows an example of such a chain of movements.

The whole activity of playing with an object can be analyzed as a whole, with a certain duration from when the person picks up the object until the person puts it down. However, one can regard this activity as a group of movements of separate anatomical components, such as wrists and fingers. In order to avoid such complex movement analysis, gestures, self-adaptors and object adaptors are considered here in a categorical fashion and will be analyzed as one single event. For instance, in the studies on boredom, the identified behaviors are also presented in a categorical fashion: doodling, playing games on the
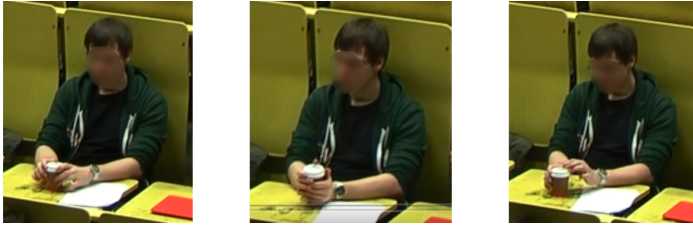
Figure 4.3. A person playing with a cup - snapshot from the observation video

mobile phone, engaging in conversations, leaving mid class (Mann & Robinson, 2009).

Other behaviors such as those that mark immediacy can be far simpler and involve sometimes only one movement, such as for instance "leaning forward". This implies a person shifting their weight from the back to the front and leaning their torso towards the table or desk in front of them. At discrete level this movement involves perhaps also a movement of the shoulders or a slight turn or torso rotation. Compared to playing with a coffee cup which implies grabbing, holding, and turning the cup in various angles through intricate wrist and finger movements, leaning forward is far easier to describe because it involves one compact movement which is easy to spot.

These differences in level of behavior description complexity can be traced back to differences in research methodology. A study on boredom required students to write down what they usually did when they were bored in class and the results are a list of categorical behaviors (Mann & Robinson, 2009). A study on immediacy by Andersen, Andersen, and Jensen (1979) proposed immediacy as Gestalt concept made of several variables that have to be considered together in the analysis of immediacy (e.g., gaze and posture together). This is an important argument for the method used in the current study as well, because it treats separate body movements and postures as unitary groups of simultaneous behaviors.

4.2.8   Nonverbal behavior multimodality

One aspect that is of interest for the virtual humans design for the virtual public speaking and presentation application is the order in which virtual humans display behaviors at a certain time. What nonverbal behaviors are simultaneous? Which ones start fist, which ones last longer and what is a coded unit?

Simultaneity or multimodality of behaviors has been studied with regard to emotions. Darwin (2000) used to describe emotions with the help of various body movements and postures that included the head, hands, arms, legs (e.g., joy–dancing for joy, clapping of hands, stamping, while laughing) (p. 196). Wallbott (1998) identified six different modalities that act simultaneous in emotion display: upper body, shoulders, head, arms, and hands. Cohn et al. (2004) measured the co-occurrence of action units eye and head motion.

In the virtual reality field, embodied conversational agents were also designed to accommodate several modalities in emotional display, such as facial action, gaze, and hand behavior (Pelachaud, 2009), which were annotated in the same time frame. Other multimodal behaviors implemented for embodied communication agents were gestural expressivity through various parameters (e.g., spatial and temporal extent or fluidity), and blended facial expressions which either mask or superpose one another to create various emotional expressions (Martin, Niewiadomsky, Devillers, Buisine, & Pelachaud, 2006).

In the case of attentiveness and inattentiveness, behaviors have been described in categorical terms (e.g., eye gaze towards someone and a torso leaned forward means attentiveness and immediacy). Yet, these behaviors have a structure and follow synchronicity rules on which modalities co-occur. The present study wishes to structure the occurring manifestations of attentiveness and inattentiveness in a real audience and to establish the synchronicity of several body modalities (e.g. head, torso, and hands). This should serve the creation of animations based on clear descriptions of behaviors and of the time patterns of their occurrence. Further studies should then test whether the implemented attentiveness and inattentiveness behaviors in virtual audiences are interpreted as such by users.

4.2.9   Summary

The theoretical background given here reflects once more the first level of the audience concept model, by delving into the potential human-like characteristics embodied by the future virtual audience.   The modalities described here represent in fine details people's postures and movements and this level of detail goes beyond the generalist nature of the traits proposed in the audience research model.   The chapter introduced definitions for audiences and their behaviors, the literature that specifically addresses nonverbal attentiveness and inattentiveness, as well as other associated behaviors such as nonverbal markers for immediacy and boredom.   Audience members can act independently of each other or they can influence each other and mimic each other's behaviors.   Both these types of behaviors will be considered in the study.   Moreover, listeners will also be assigned roles.   There are those who act voluntarily and there are those who are distracted by others and, due to that, respond to external stimuli, such as in the case of involuntary attention.

Not all described behaviors could be assigned to attentiveness or inattentiveness in the absence of gaze direction.   However, the multimodal approach of the research allows all behaviors to be analyzed in combination with a eye gaze and therefore carry a meaning of either attentiveness or inattentiveness.

CHAPTER 4.3

Methodology

The study presented here is based on a similar research design as the previous ones. The goal was to reveal how people (in this case university students) behave in a real audience during a lecture, in order to transfer these behaviors to virtual audiences. The study is exploratory and descriptive (de Vaus, 2001). In order to acquire data about audience behaviors, a participatory observation was conducted. This method has been previously used in other studies for similar purposes (e.g., Kang et al. (2013) and Poeschl and Doering (2012)). This method is the most fitting for the particular purpose of this study because it permits a thorough quantification of the data (the observed behaviors). Through the exploratory research design, the collected video data permitted a thorough analysis of behaviors beyond the two research questions. This is visible in the results section where the data analysis extends to more behavior characteristics than the ones inquired initially.

4.3.1 Research questions

The study focused (1) on audience reactions towards the speaker such as attentiveness and inattentiveness, and (2) on one type of group behavior–observable audience responses. The first research question of the study was RQ1: *How does attentiveness and inattentiveness manifest nonverbally within an audience?* Nonverbal behaviors of listeners are coded and then grouped based on criteria that differentiate between attentiveness and inattentiveness. These criteria will be explained in the codebook development section.

The second research question was RQ2: *What nonverbal behaviors in the audience trigger nonverbal responses from fellow audience members at group level and how do such behaviors spread?* The answer will help complement the interview findings on group dynamic with empirically-founded descriptions of such behaviors.

4.3.2   Codebook and code sheet development

This section discusses the coding methods existing in literature and how they have been implemented to develop the codebook and the code sheet for the study.

4.3.2.1   Coding methods

Nonverbal behaviors can be annotated in various ways, and there are generally two coding methods. The first one codes movement features (e.g., effort, shape, and occupied space, like in Labanotation (von Laban, 1980) or in the Body Action and Posture (BAP) coding system (Dael, Mortillaro, & Scherer, 2012)). The other one codes movement type (Dael et al., 2012). The Body Action and Posture coding system is a descriptive notation system that contains 141 behavior variables to code emotion expressions (Dael et al., 2012). At anatomical articulation level, the coding system covers head, trunk, and arms. At feature level, the code describes how coded body parts are moving, such as direction–left or right). At functional level, the system includes gestures: emblems, illustrators, and manipulators (Dael et al., 2012).

A similar coding method was implied by Kang et al. (2013) who used a codebook to identify head, gaze, arms and hands, torso, and leg postures within a real audience in order to create behavior combinations for virtual humans. The code list provided detailed positions of arms and hands, as well as directions of movements for the torso. However, the researchers didn't focus on identifying a specific behavior in the analysis, but provided behavior combinations that were later analyzed by subjects who evaluated the virtual audiences.

The second analysis method uses a theory-based variable construction (Babad, as cited in Harrigan, Rosenthal, and Scherer (2008)), which classifies body movements into typologies adapted usually to the purpose of the research, such as devising coding systems for the study of particular attitudes or emotions. Such systems are tailored for specific research purposes and are customized to address specific body movements, postures, and gestures relevant for the identification of the particular researched attitude (Dael et al., 2012).

The current observational study combines elements from both coding methods. Body segments are coded in a categorical fashion adapted to the purpose of this study. Eye gaze and head are segments used in

identifying attentiveness and inattentiveness behaviors. The *features* of these movements such as direction, strength, or speed are simplified to adapt the particular needs of the research and refer only to the two directions: front and left/right. In the codebook, directions are grouped, at times, to simplify the coding process. For instance, there is only one code for averted eye gaze and/or head position, independently of the side of the head axis–right or left. The original codebook contained a left/right separation, but this was dropped after the pretest. Only two features for eye gaze remained: averted (which contains left/right/down) and direct (frontal).

4.3.2.2   Coding categories

The codebook contains categories for all relevant movements that can be associated with attentiveness, inattentiveness, and observable audience response. The choice of categories was based on empirical studies on nonverbal markers of these three types of studied behaviors, as well as on a pretest observation of a seminar at Technische Universität Ilmenau. The recording of the pretest audience comprises 80 minutes, out of which only 10 minutes were coded. In the pretest, behaviors were coded inductively and served as staring point for the codebook of this study.

Three main groups of behavior categories emerged from literature review and the pretest.

- Basic behavior elements. These behavior elements contain eye gaze and limb positions and postures. They are called basic elements because they can be regarded as the basic unit in body posture and movement analysis (Coulson, 2004). Limb positions and postures considered here are: head, arm and hand, and torso. Leg and foot movements were left out because they were not visible in the video recording due to the desk structure which obstructed the view of the lower part of the body from waist on.

  These basic positions and postures are part of the following codebook categories: eye gaze and/or head behavior; agreement or disagreement behaviors; specific head positions; head and arm positions (static); body (torso) orientation postures (static). Conversation (speech) was also assigned to the basic behavior category.

Head and eye gaze were coded together in all situations, since
the distance from the camera to the observed persons was far
too big to notice when the eye gaze was oriented in a different
direction than the head.   Therefore, the head direction also
gave the direction for the gaze, in case the eye gaze was not
clear. Apart from this, nodding and head shaking which denote
agreement and disagreement were further included (Ekman, 1979;
Richmond, 2003).

The postures chosen for the codebook reflect the upper body seg-
ments proposed by Coulson (2004): neck, both shoulders, chest,
abdomen, upper arms, and forearms. Two body directions were
proposed for body postures, in accordance with the movement
features approach to coding (Dael et al., 2012): "averted" and
"front-facing posture", differentiating between a torso oriented to-
wards the front of the lecture hall, and a twisted torso to left or
right. Other coded postures and positions included among others
"crossed arms" and "one or both hands under the table".

- Complex behavior elements. These behaviors consist of complex
  movements that are easily described in a categorical way, as
  unit of behavior rather than be interpreted through single basic
  behavior clustering.   For instance "playing with an object"
  was coded as a single complex behavior element unit instead
  of describing the exact movements of the hand while playing.
  Many complex behaviors used in the codebook were taken from
  literature and from the pretest observation and compiled into a
  single common list.

  Complex behavior elements are part of the codebook category
  "complex body movements" and comprise: leaving before class
  is over; receiving something from a neighbor; giving something
  to a neighbor; playing around with objects; ordering objects;
  typing (or actuating) on a mobile device or on laptop or any
  other mobile device; taking notes; drinking or eating; unpacking
  objects or packing up; stretching; yawning; coughing or sneezing;
  self-grooming or scratching; gesturing in a conversation; moving
  with low visibility (usually under the desk); and other.

- Trigger elements presented in Figure 4.4 represent a separate cat-
  egory that helps differentiate between attentiveness and inatten-
  tiveness.   Furthermore, this category was used to differentiate

between types of inattentiveness: (1) the deliberate inattentiveness when a person decides to look away from the speaker and engage in an activity unrelated to the lecture and (2) the induced inattentiveness experienced by a person who is disturbed by someone in their vicinity. The graph below depicts the three variables included in this category:

1. the code for "other" (fellow student) represents the case when the coded person reacts to what someone around them does (comes in late, asks to borrow a pen, etc.) (induced inattentiveness);

2. the code for "self" represents a person who is the active agent of the displayed behavior (deliberate inattentiveness);

3. the code "speaker/lecturer" is the default trigger code assigned to the coded person when they are looking towards the lecturer (attentiveness).
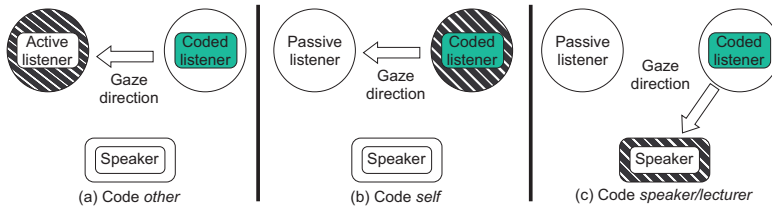
Figure 4.4. Attentiveness and inattentiveness triggers

Facial expressions, such as smiles and laughter and the coding of any emotions were left out after a pretest of the codebook. They were hard to identify due to the distance between audience and camera and due to the filming angle. Furthermore, due to the modular nature of behavior coding in this study, it is expected to see a large number of various behaviors made of components that get to assemble and reassemble with each other in new ways and build new behavior configurations throughout the observation process.

4.3.2.3   Time line and base position

To ease the animation process, each observed and coded behavior received a starting and an ending point. This is meant to help animators with time-related behavior characteristics. Start and end

points of observed behaviors were recorded in seconds in the coding platform, which was developed particularly for this study. A time line was used to count the frequencies of behaviors and their duration.

Another aspect of human behavior coding is the base position which is used in order to offer a starting point for the features of the coded movements, such as rotation angles or direction (Dael et al., 2012; Gifford, 1994b). In the present study, coding was made dynamically from the last recorded position to the next one. This was particularly helpful when reading the code lines, because one can observe in the final coding table how people switch from one movement to the other, without having to go through and code the base position every time.

4.3.3 Sample

A group of 38 media and communications students from Technische Universtät Ilmenau were recorded during a 40-minutes lecture (32 female and 6 male). Students were attending a lecture in the Bachelor course in the winter semester 2012/13. The students were dispersed across all the lecture hall. Two students decided in the last minute not to participate and were seated in a part of the lecture hall that was not being filmed. Out of the 38 participants, only 37 were eligible for coding, as one person was seated in the the extreme corners of the filming frame and didn't fit completely within.

4.3.4 Data collection

Three cameras were used to record the lecture. All were positioned in the front upper corners of the hall, behind the lecturer and facing the students. Due to the construction of the lecture hall, it was necessary to orient each camera differently, to ensure that it covers as much as possible of the sitting area. Two cameras were positioned on small balconies on both right and left flanks of the front wall of the lecture hall. These cameras captured the students from an angle above their heads. A third one was positioned approximately at the same level with the standing lecturer on the right flank of the lecture hall. This angle captured the students from a low angle and helped record those students who were not visible through the upper cameras.

The researcher also establish a blind spot in the lecture hall, where students could sit if they decided not to participate to the study. Out of all the present students, two chose to move to the blind spot and the other 38 remained seated in their original positions and returned the signed agreement of participation. As previously stated, no self-report questionnaires or knowledge tests were administered at the end of the lecture, therefore only *displayed* markers of the coded behaviors will be reported in the results chapter.

4.3.5   Coding interface

To accommodate the multimodality factor of coded behaviors, the coding occurred with a specially designed interface which allowed the researcher to play and stop the recording and introduce the codes of observed behaviors in the selected time frame. Figure 4.5 shows a screenshot of the used interface. Each code was assigned a separate button in the interface. The coding started with the "trigger" and "head/gaze direction" categories, with the latter also recording automatically the duration of the coded time frame. After all other buttons corresponding to the observed behavior were checked, the researcher would press the "Entry" button which would save into a database the fully observed behavior combination for the current time frame.

Figure 4.5. Screenshot of the coding interface

The time frame was created automatically every time the play and stop buttons were pressed and it was as well saved in a database. For the coding procedure, several rules were developed and they are explained in the following section.

4.3.6   Coding rules

Nonverbal behaviors can be coded against a timeline, for instance every five seconds (Gifford, 1994b). In studies on attention during class, frequency of note taking was counted in 10-minutes (Hartley & Cameron, as cited in Wilson and Korn (2007)) and five-minutes intervals (Maddox & Hoole, as cited in Wilson and Korn (2007)). In a study on deceptive nonverbal behaviors, frequencies of movements were counted for the whole length of the studied time (between 20-30 seconds in the study) (Vrij, Semin, & Bull, 1996). For the present observation, behaviors were analyzed in fixed time segments, at the beginning (five minutes), middle (three minutes), and end (two minutes) of the recorded lecture. It was decided to allocate more time at the beginning of the lecture than at the end because people could be performing more of the expected movements in the codebook, such as unpacking objects and settling in. Towards the end of the lecture it was expected to identify behaviors such as leaving before the formal ending of the lecture or packing up.

The coding procedure followed several rules, such as coding simultaneity, coded unit, and discriminatory behavior.

After the coding of the recorded material, a matrix occurred with a column containing the noted times of behavior occurrence and rows containing co-occurring behaviors, as seen in Figure 4.6.

Each row shows *a behavior combination* that includes variables from all codebook categories. Behavior combinations can have a unique fingerprint, which means that they contain a unique mixture of identified variables for each of the nonverbal modalities researched here. When behavior combinations with the same unique mixture occur several times, they are counted as one type of unique mixture and are named *patterns* due to their repetitive nature within the corpus. To identify unique combinations and patterns of combinations in the data analysis, a script was developed in MATLAB.

Duration	Participant	Action trigger	Eye gaze	Agreement/ disagreement	Head pos.	Hand and arm pos.	Body posture	Body movements	Conversation
2	707	2	5	0	11	20	21	36	39
8	707	3	6	0	11	19	21	29	0
3	707	1	4	0	11	20	21	37	0
6	707	2	5	0	11	20	21	37	39
3	707	3	6	0	11	20	21	31	0
2	707	1	4	0	11	20	21	38	0
4	707	2	5	0	11	20	21	38	39
6	707	1	4	0	13	19	21	0	0
2	707	2	5	0	11	19	21	35	0
1	707	1	4	0	13	19	21	0	0
7	707	2	5	0	13	19	21	0	39
34	707	1	4	0	13	19	21	0	0
16	707	3	6	0	11	20	21	38	0
1	707	1	4	0	11	17	21	0	0
3	707	3	6	0	13	19	21	0	0

Figure 4.6. Code sheet excerpt

4.3.7 Inter-rater agreement

An inter-rater agreement test was performed using a fixed time frame
of five seconds, in which each coder would have to mark the occurrence
of each of the behaviors given in the codebook (Gifford, 1994b). Two
coders observed two people in the audience in a 10-minute time interval
(besides the researcher, another person was trained and assigned to
code). The agreement was computed with Cohen's Kappa (J. Cohen,
1960) and the overall coefficient was 0.85, revealing a high level of
agreement, close to the one achieved in a similar study on nonverbal
audience behavior (Kang et al., 2013). Table 4.1 presents the coefficient
for each category.

The lowest calculated value was for action trigger, since people
performed a lot of head movements and it was difficult at times to
notice whether they are distracted by someone else or they change their
gaze direction voluntarily. For gaze, agreement was high (0.82), and
this was a reassurance that coders could differentiate between frontal
and averted gaze. Furthermore, this was also reassuring that gaze was
a sensible choice in discriminating between attentive and inattentive
behaviors in the data analysis.

Table 4.1

Inter-rater agreement (Cohen's Kappa)

Category	Cohen's Kappa
Action trigger	0.7
Gaze/head direction	0.82
Agreement/disagreement	1
Head position	0.99
Hand and arm	0.76
Body posture	1
Complex body movements	0.73
Conversation	0.84

4.3.8   Ethical considerations

The observational study was conducted in agreement with the American Psychological Association principles of ethics (D. Smith, 2003). Credit is given as follows: the study was conceived and partially conducted together with a Bachelor seminar group at the Institute of Media and Communication Sciences, at TU Ilmenau, in the winter semester of 2012/13. Students participated to the development of the initial version of the codebook and to the organization of the first pretest. A group from the seminar participated together with the researcher to the filming of the student audience.

For the participant observation, the researcher obtained verbal consent from the lecturer to present the study and to brief the students about the observation a week before the filming. On the filming day, students were given a written description with the scope of the study and they signed a written consent of participation. No other personal data was inquired. They were informed about the data confidentiality policy–to use the images for research purposes only and always with blurred faces. The contact data of the researcher was also provided for any further questions. Participation was voluntary with no financial incentives.

The coding platform was developed with the help of Mr. Sherief Emam from the Department of Economic Policy at TU Ilmenau, and data analysis algorithms in MATLAB were developed in collaboration with Mr. Bogdan Barbu from the Department of Electrical Engineering at TU Ilmenau.

CHAPTER 4.4

Results

The following chapter presents findings on manifest behaviors (Compton et al., 2012) that show attentiveness and inattentiveness, as they are displayed by the student audience during the recorded lecture. These results give an insight into the synchronous manifestation of nonverbal modalities, which can serve in turn as animation options for multimodal virtual humans. The findings are based on visual observation only and do not include any analysis of audience vocal behaviors, lecture content, or of lecturer behavior.

Out of the 37 observed students, 14 students were randomly chosen to be coded for a total duration of 10 minutes: three minutes at the beginning, five in the middle, and two minutes at the end of the 40-minute long lecture. Based on the coding interface, it was possible to code all behaviors that co-occur simultaneously at any selected point in time. A coding line containing several simultaneous behaviors is called *behavior combination*. When a behavior combinations appears only once, it is called a *singular behavior combination*. If a behavior combination occurs at least two times in the coded corpus, it is called *behavior pattern*. To ensure that all identified behaviors are logical and do not contain contradictory terms (e.g., coding "no hand movements" together with "ordering objects"), a thorough search of all data was performed. All illogical or suspect combinations were removed or corrected. A total of 17 suspect combinations were removed from the initial corpus before the begin of the data analysis.

In total here were 966 observed behavior combinations. Some of them occurred once, some occurred several times and became patterns. A thorough search was conducted in the corpus of coded behaviors and all coded combinations were counted. These 966 coded combinations could be reduced to 225 *unique* behaviors, each with its own unique fingerprint. To put these numbers into perspective, a script calculated the total of all possible and logical combinations including all body

modalities based on the codebook, and there are around 75.000 possible unique combinations or different ways that people could sit and move while listening to the speaker.

The fact that all 966 coded combinations can be boiled down to a rather manageable number of 225 unique behaviors proved that many behaviors were repetitive, which means that once in a while people displayed the same movement combination all over again, for instance, looking to the speaker with hands under the desk without making any other movement.

There is a major advantage to this high number of diverse behaviors: it grasps the great diversity of human movements and it offers a very large pool of potential behaviors that can be animated, therefore the risk of highly repetitive behaviors can be avoided. Previous studies on audiences have used a few behaviors to display positive and negative behaviors, and the study here stays as proof that, in fact, people behave in much more diverse fashion. This diversity also emerged in the expert interviews as key factor for a virtual audience, where speakers would have to learn to interpret the nonverbal signals of the audience and adjust their content and delivery style accordingly.

The results chapter followed the research question on attentiveness and inattentiveness and reports further findings on behavior activation. Next, the group dynamic is discussed. Due to the rich corpus, further analyses were conducted to expand the data assigned to the research questions and address other features of behaviors identified in the corpus, based on their order of occurrence: synchronous (parallel behaviors) and sequential behaviors. Another subchapter deals with behavior duration. The last part is dedicated to the concept of saturation area, which explains the coding potential of the used method.

4.4.1   Attentiveness and inattentiveness

In spite of the presence of the cameras, students didn't seem distracted by them, except for a few times at the beginning of the lecture. They seemed to ignore the cameras, in accordance to the wear-off effect identified in other studies (Weick, as cited in Wiemann (1981)). Moreover, Wiemann (1981) demonstrated that video recording has no effect on behaviors out of conscious control, such as head nods, gestures,

or amount of gaze directed towards the other. Therefore, it is safe to say that coded behaviors can be attributed to the interactions among speaker and students, and not to the intrusion of the recording cameras.

The research question posed here was: *How does attentiveness and inattentiveness manifest nonverbally within an audience?* The analysis was exploratory and it wished to identify the behaviors that people display when they pay attention to the speaker or when they are busy with other activities.

4.4.1.1 Attentiveness

The convention for the current analysis was that eye gaze direction is the discriminating behavior that differentiates between attentiveness and inattentiveness. Gaze and implicitly head direction towards the speaker is a pervasive behavior in the literature on attentiveness and immediacy (Frischen et al., 2007; Hale & Burgoon, 1984; Norton & Pettegrew, 1979). Therefore, in the data interpretation, any behavior that contained the gaze and/or head oriented towards the front where the lecturer was positioned was assigned to attentiveness.

Results show that this behavior is more than just sitting totally inactive and looking at the speaker. It contains behaviors where body postures and hands are actively involved. People rarely sit still and do nothing, even when looking at the lecturer. In total, attentiveness manifested in 47 different ways. In contrast, inattentiveness manifested in almost four times the number, because more behaviors from the same category could be combined simultaneously, as explained in the next section.

The majority of coded attentiveness occurred only once (21 unique combinations). What is interesting about them is that, even if people displayed an attentive posture, by looking at the lecturer and having a front-facing posture, they would perform various activities such as conversing, taking notes, actuating mobile devices, or bending under the desk to look for things in their bags as shown in Figure 4.7

There are several behaviors that occurred repeatedly, at least two times (26 unique patterns). The most common display of all identified behaviors (72 times) is of a person who sits still and watches the lecturer. Other attentiveness patterns represented people who would stand rather still except for scratching or grooming themselves (using self-adaptors), eat, or drink.
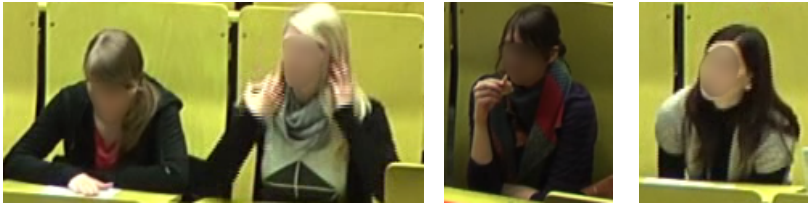
Figure 4.7. Display of attentive behaviors

In literature, body orientation towards a speaker has been considered as sign of attention (Norton & Pettegrew, 1979). According to one of the interviewed training experts (RM), the more is the case for a public speaking situation, where it is expected that listeners pay attention to the speaker and respect the rules of the speech setting. However, as these findings show, people are active in spite of always facing the speaker. This is the point where one can observe differences between what one could expect under "attentive" and how that behavior is actually played out. Even if people watch the speaker, they prove rather versatile at shifting their posture as well as performing various movements with their hands.

A first conclusion is that a frontal gaze/head posture appears with both basic behavior elements and complex ones and Table 4.2 centralizes some examples of behaviors people display while looking at the speaker.

Apart from these attentiveness markers, there was another group of behaviors that can be assigned to attentiveness and which contains shifts in gaze direction. Several students were observed taking notes and repeating a certain behavior sequence: looking to the front (to the lecturer or to the presentation), looking to the desk (to their notes or notebook), and then looking back to the lecturer. Literature in the education field proposed note taking as a sign of displayed attention (Wilson & Korn, 2007). Based on this consideration, the sequence of looking at the lecturer and then to the desk, accompanied by writing down was considered a marker of attention. While people perform these sequences of alternate behaviors, the rest of the body is also in low activation mode: few or no posture movements, very little conversations, and hands are both still, except for the writing activity. For the virtual audience, note taking on paper can be complemented by note taking on a laptop or on a tablet.

Table 4.2

Examples of nonverbal displays of attentiveness

Gaze/head direction	Basic behavior elements	Complex behavior elements
to the front	both hands on/under the desk; holding objects (cups, phones, pens); resting the head on the chin; crossing the arms; torso towards the front.	playing with objects; ordering objects on the desk (papers, bags with food, phones, tablets); scratching (the nose, the face, the hair) and/or grooming (arranging the glasses, the hair, cleaning the clothing); taking off coats, hats; putting bags on the desk/on the chair/on the floor; unpacking objects (sandwiches, laptops); actuating the phone; taking notes.

4.4.1.2 Inattentiveness

The major marker for inattentiveness in literature is the averted eye gaze direction (MacLean et al., 2009; Remington et al., 1992), which is sometimes also accompanied by various distracting activities, such as doodling, playing games on the mobile phone, leaving mid class, etc. (Mann & Robinson, 2009).

According to the trigger factors discussed in the methodological chapter, a person can become inattentive either when someone around distracts them or when they deliberately look away from the speaker. In the codebook, based on the pretest and on reviewed literature, an inattentive person could: look down to the desk, to the left, or to the right, could daydream, sleep, and look around. All these variables could co-occur with all other basic and complex behaviors in the codebook.

In total, inattentiveness manifested in 178 different ways. This means that for every displayed attentive behavior, there were almost four others which displayed inattentive ones. People can perform many
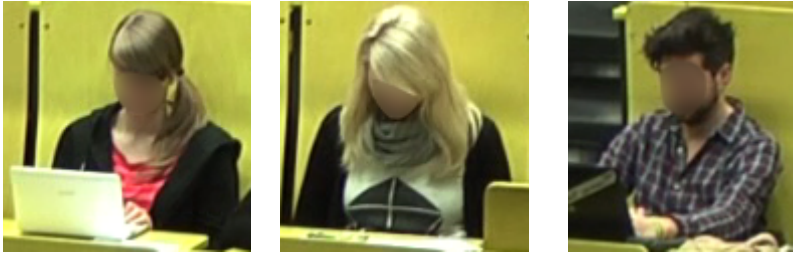
Figure 4.8. Display of inattentive behaviors

body movements and engage in conversations that require changes in gaze direction, they can actuate mobile devices or laptops, pass on objects, engage in conversations, or show objects to fellow students.

One particularity of inattentive behaviors was that they manifest through combinations that engage many modalities (head, hands and arms, torso). Most of combinations that engage variables from six or seven different categories represent inattentiveness, which means that these are highly active behaviors. The most frequent inattentiveness behavior that was coded in the corpus (35 times) represented a person who is looking at the desk, with a front-facing posture, and who is busy actuating mobile devices, such as a tablet or a mobile phone (Figure 4.8).

Table 4.3 exemplifies some of the behavior combinations for inattentiveness.

Apparently, comparing Table 4.2 and Table 4.3, it can be concluded that attentive and inattentive people perform in general the same movements and only gaze direction differentiates them. The section on behavior activation addresses this aspect in more detail.

The majority of the coded inattentiveness behaviors occurred less than 10 times, with a long tail of singular unique occurrences. Among those rare behaviors, are those depicting people who perform rather counterintuitive actions, such as conversing while typing on their phones or tablets.

In spite of their large number, results from inattentiveness represent a recalculation after the codebook was changed following a first inter-rater agreement test. Originally, all averted gaze behaviors were

Table 4.3

Examples of nonverbal display of inattentiveness

Gaze/head direction	Basic behavior elements	Complex behavior elements
looking down to desk; looking left or right; daydreaming; sleeping; looking around.	both hands on/under the desk; holding objects (cups, phones, pens); crossing the arms; averted torso.	playing with objects; ordering objects on the desk (papers, bags with food, phones, tablets); scratching (the nose, the face, the hair) and/or grooming (arranging the glasses, the hair, cleaning the clothing); taking off coats, hats; putting bags on the desk/on the chair/on the floor; unpacking objects (sandwiches, laptops); actuating the phone; taking notes.

divided between left and right directions, and front-facing postures were divided between leaned forward and leaned backwards. This detailed description of direction yielded more single unique behaviors, but prevented a proper interpretation of data. Merging directions led to a more manageable list of findings. In the design process of virtual humans, directions can be expanded back to the left-right and up-down orientations.

Another finding on inattentiveness refers to involuntary attention away from the speaker. Involuntary attention was defined in the theoretical chapter as the shift of attention from a stimuli towards another one that enters the visual or auditory field of a person (Eimer et al., 1996). Around a third of all inattentiveness behaviors were triggered by the intervention of a fellow student who either initiated a conversation, passed on an object, showed something to the coded person or simply performed a movement that distracted the attention of the coded person. Among the distracting movements were unpacking

objects, ordering objects, and entering or leaving the lecture hall. By comparison, deliberately inattentive people chose to look somewhere else than to the speaker, without being distracted by anyone. However, for virtual human behavior design, such inattentive behaviors can be used in turn as distraction for other members of the virtual audience. This finding offers an empirically-grounded suggestion about the distribution of triggers and of inattentive behaviors within a virtual audience: most of the times, people chose to look somewhere else, and distractions from fellow students occurred in a ratio of 1(involuntary):4(deliberate) inattentiveness. Of course, the findings mirror this specific observed audience, and in order to be able to generalize them to any audience, a study with a bigger sample should follow.

4.4.1.3 Summary

Attentiveness and inattentiveness were calculated by recording eye gaze direction and every other behavior a person displays while having a certain eye gaze direction. Findings showed that, while sitting and listening to a speaker, people can multitask and perform a lot of body movements and hand movements, even if they seem attentive. Also, they may as well stay still but look somewhere else, which can as well denote that they aren't paying attention.

Most of the observed behaviors represented inattentiveness, with a ratio of 1(attentiveness):4(inattentiveness). Most inattentive behaviors were unique. This could be explained by the fact that people can look in many directions away from the speaker and that each of these gaze directions can combine with other variable, giving a very large number of possible combinations.

Coding action trigger elements for inattentiveness proved to be a good strategy, because it allowed to map who *initiates* behaviors and who *follows* them. Based on this principle, the virtual humans in the audience can be assigned roles of "initiators" or "followers", which can in turn justify interaction among audience members. Moreover, it was also empirically proven that people tend to perform less movements when they look to the front, and to the speaker, than when their gaze is oriented elsewhere.

4.4.2   Activation

Data analysis showed that both attentive and inattentive behaviors can contain static behaviors and movements. In order to see whether movements are correlated with gaze direction, Chi-Square tests were performed for all 966 coded combinations.   Gaze was chosen as independent variable and body movement as dependent variable.

Table 4.4

Correlation between body movement activation and gaze direction

| | | Gaze direction | |
		Frontal (n=280)	Averted (n=686)
Body activation	Movement (n=554)	43%	63%
	No movement (n=412)	56%	37%

The results presented in Table 4.4 show that there is a significant correlation between gaze direction and body activation, with $\chi^2_{(1)} = 29.037$, $p < .001$. However, the effect is rather small ($\phi = -0.173$, $p < .001$). The contingency table shows that, when people don't look to the speaker (averted gaze), they perform more movements compared to when they look to the front in the direction of the lecturer. For instance, when they look at the desk, they usually engage in manipulating various objects, actuate mobile devices, or pack and unpack objects. When they display eye gaze attentive behavior, they move their bodies less compared to when they look elsewhere.

This information gives a few simple rules for animating virtual humans and ultimately obtaining various attentive and inattentive audiences. When a virtual character is looking towards the speaker, it could be less active than when it looks elsewhere, hence fewer body movements animations could be created. When a virtual character is looking to the desk or to a colleague, it could perform various hand and body movements (e.g., order objects, actuate phones, pass on objects, etc).

4.4.3   Group dynamic

The second research question of the study investigated what nonverbal
behaviors in the audience would trigger nonverbal responses from
fellow audience members, at group level. In literature, behaviors that
occur when people observe the action of the others and start behaving
similarly are called observable audience responses (Hylton, 1971).

One type of behavior that could be observed to occur and spread across
the audience was involuntary attention as response to someone entering
the lecture hall late. Any student who entered the room late or left
early attracted the gaze of fellow students in an orderly fashion. As the
student would climb the stairs to leave the room, the students seated in
the front would be the first ones to look at him/her, then the students
behind these rows would follow. This pattern would repeat until the
person leaving would enter the visual field of students from all rows up
to the door. The Figure 4.9 illustrates in a time-lapse of 13 seconds this
effect on a selected area of the recorded audience. The red dot marks
the persons in the audience whose behavior was affected by the leaving
person.

Figure 4.9. Example of involuntary attention that spreads within the
audience when a person leaves before class is over

Another such behavior, typical for German universities (Beiseler, 2004)
was knocking on the desks at the end of the lecture. Knocking on
desk was identified in the observed audience, at minute 41:05, after
the lecturer announced that the class will recess in a few minutes, and

lasted for about four seconds. On both sides of the lecture hall knocking started simultaneously for a group of five-seven people, followed in the next second by almost the double. Among the whole audience, there were also people who didn't knock at all.

4.4.4 Further characteristics of attentive and inattentive behaviors

This subchapter contains further analyses of the whole coded corpus and presents interesting results with regard to the observed manifest nonverbal behaviors, independently of their meaning of attentiveness or inattentiveness. These findings are relevant from a design perspective, as they help structure potential animations in audience scenarios.

The first analysis targets behavior complexity–how complex are nonverbal displays. The results are meant to help structure and prioritize the animation of virtual humans, by establishing an average complexity level for all behaviors. Even if behaviors are categorical (e.g., a person is playing with a pen), people do not always use the same body modalities to express themselves, some play with a pen while looking to the speaker, some while reading their notes. Therefore, in order to achieve behavior variety, it is necessary to look at how complex such behaviors are. As a brief reminder, behaviors were coded in a dynamic fashion, on a timeline given by the duration of a gaze direction. All behaviors that were displayed simultaneously with the coded eye gaze/head direction were coded on the same line in the code sheet. Unlike in the BAP (Dael et al., 2012) or the SKANS V (Seated Kinesic Activity Notation System) (Gifford, 1994b) codebooks where behaviors are coded separately from one another, the coding here focused specifically on how behaviors combine with one another. The first analysis revealed that, among all the hundreds of coded behaviors, behaviors can be grouped into two categories: (1) the synchronous or co-occurring ones, called *parallel combinations* and (2) behavior combinations that follow one another in an orderly fashion which were called *sequential combinations*.

The second analysis targets behavior duration–what is the length of coded behaviors? Both attentiveness and inattentiveness manifested in highly diverse ways and with diverse durations. The analysis shows that such behaviors can last from one second to several minutes in a row.

The third analysis tries to identify a potential saturation area where adding new persons or new time segments to the analysis would not enrich the corpus with new behaviors, but would enlarge the frequencies of the already coded ones. This analysis was meant to help plan similar studies, by deciding on how to start the analysis and on how much coding effort is needed to obtain behavior variety.

4.4.4.1   Parallel behavior combinations

Several categories had to always be coded, such as action trigger (speaker, self, or other), eye gaze/head direction, head posture, and body posture. These elements would always be possible to observe for each person in the audience. Action trigger and gaze differentiate between attentiveness and inattentiveness, whereas the other ones are features of postures (e.g., the torso orientation of a person). The remaining other four categories (agreement/disagreement, hand and arm behaviors, complex body movements and conversations) represent movements that are not always coded, as sometimes people choose to sit completely still. As a result, the minimum number of categories which could be simultaneously coded is four–representing a person is who is only looking in a certain direction (either to the speaker or somewhere else), without performing any movements at all. The maximum possible coded ones depicted a very active person involved in conversations, who performed various head, hand, and body movements (coded as inattentive).

The first finding about the behaviors displayed by audience members is that, in the coded corpus, they have a normal distribution and peak at six simultaneous behaviors in a combination. Considering that four categories have to be coded all the time and are static, most of the actions that people performed (107 unique combinations or almost 50%) covered only two more categories, out of four more possible. Table 4.5 presents the number of unique combinations identified in the corpus for all number of categories in the codebook. In the last column there are the total frequencies of the occurring behavior combinations.

There aren't many unique behavior combinations appearing at the extremes. As mentioned above, the most frequent ones have only six coded categories. They are followed by the behavior combinations where seven categories have been coded, which are with almost 50% less numerous (63). The fact that there are so few behavior combinations where only the first four categories are coded (6 unique combinations

Table 4.5
Total behaviors and their frequencies

No. of coded categories	No. of unique combinations	Total number of occurrence
8	8	9
7	63	144
6	107	481
5	41	279
4	6	53
	225	**966**

in total) proves that people perform various movements when they sit, and rarely simply stay inert at the desk while listening to the speaker.

This is an important finding with regard to how behavior complexity could be a factor for prioritizing which behaviors are possibly more likely to occur in an audience. Very active behaviors, where all eight categories are coded as well as very still ones (four coded categories) were, in the case of this audience, the exception. Most of observed behaviors, either for attentiveness or inattentiveness contained six different types of nonverbal behaviors. The findings should not be generalized across other audiences due to the small coded sample, but they can give a direction for behavior animation in a virtual audience for public speaking training as well as for future research on behavior complexity in various audience types, across other speech settings (e.g., conferences, business presentations, workshops, etc.).

Another calculation can reveal how many times did behaviors repeat throughout the corpus. This aspect would be relevant especially for the case of short training scenarios. Public speaking experts (CS, RM) in the first study stated that presentations in training sessions can take from 1-8 minutes. The sample analyzed here contains a ten-minute time frame. For this time frame, most behavior combinations (119) manifested only once (see Table 4.6).

The data presented in Table 4.6 can be interpreted as a sign of high heterogeneity among coded behaviors. This is good news for anyone who wishes to design diverse behaviors in the virtual audience, because they have a wide choice of behaviors to select from. It is possible

Table 4.6
Number of combinations reported by frequency

Type of combination	Number of unique combinations
Singular unique combinations	119
Pattern(s) that occur 2 times	30
Pattern(s) that occur 3 times	17
.....	....
Pattern(s) that occur 72 times	1
Total = 26 type of combinations	Total = 225

though that coding a longer period of time might raise the chance of repetition of these behaviors. At the other extreme, there are behaviors that repeat many more times. There is one pattern that was identified for 72 times, which represents almost 10% of the whole corpus. This pattern depicts an attentive person who is looking at the lecturer, with a body oriented towards the front, with one or both hands under the table and who otherwise is completely motionless.

The coded behavior combinations are subject to both codebook design, but also to factors external to the coded person, such as the architecture of the space they occupy. At some point, everyone puts their hand on the desk. Everyone can keep their hands under the desk and on the lap during the lecture, at some point. Everyone packs and unpacks at the beginning and end of the lecture. Such rituals may differ across settings, where people simply sit on chairs and have no desks or attend meetings where they carry no coats with them. The configuration of behaviors combinations could therefore look very differently. Figure 4.10 shows the most frequent pattern of the attentive person.
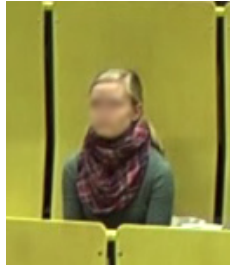
Figure 4.10. Example of the most frequent attentive behavior

An example of a pattern that occurred only once and which also covered all eight possible coding categories represents an inattentive person who is unpacking various objects from a bag under the desk, looks in that direction and also has a conversation with a fellow student (see Figure 4.11). This example was of a multitasking person who, within the context of this analysis, is very active. Below is a snapshot of this behavior.
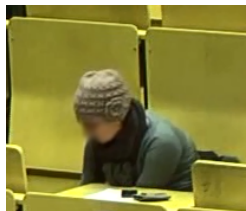
Figure 4.11. Example of singular unique behavior combination

To wrap up, the purpose of identifying parallel or co-occurring behavior combinations was to offer a database of behaviors that combine several nonverbal modalities. As the amount of coded data suggests, most behavior combinations occur only once. This could be explained by the small number of coded people and by the time frame selected for analysis. It is expected that the more people are coded or the longer the coded time, the fewer new unique behaviors occur and the ones that have been already coded keep repeating. Nevertheless, there is a high variety in what people can do simultaneously as well as how complex their behaviors are. This richness can be transferred to virtual audience with the certainty that they are empirically founded. This

behavior variety addresses the needs of the public speaking experts (CS, RM) who expect diversity within the audience that mirrors real life speech and presentation situations. Moreover, the findings on parallel behaviors offer a rich insight into concrete behaviors that involve all possible upper body movements and postures.

4.4.4.2   Sequential behaviors

The code sheet contains lines of behaviors that have the duration of the coded eye gaze. However, as explained above, behaviors from different categories can have different durations. This is why some behaviors can continue to happen also when the gaze direction has changed and a new behavior combination line has been recorded.

The analysis of sequential behaviors can help find the relation between behavior combinations and ensure that, when animated, each behavior flows naturally from one into another and in accordance with the fluidity of natural movements. Here, fluidity refers to the smoothness of a gesture and to the smooth continuity between movements (Hartmann, Mancini, & Pelachaud, 2005) and reminds of the movement analysis system developed by Laban, where flow refers to movement continuity, 'ongoingness', and progression (Lourens, van Berkel, & Barakova, 2010).

Figure 4.12 is an example taken from the code sheet of the coded corpus and it illustrates what happens to behaviors that last longer or shorter than the coded eye gaze duration. Each code was assigned a random color to mark the durations of the observed behaviors.

Duration	Participant	Action trigger	Eye gaze	Agreement/ disagreement	Head pos.	Hand and arm pos.	Body posture	Body movements	Conversation
2	707	2	5	0	11	20	21	36	39
8	707	3	6	0	11	19	21	29	0
3	707	1	4	0	11	20	21	37	0
6	707	2	5	0	11	20	21	37	39
3	707	3	6	0	11	20	21	31	0
2	707	1	4	0	11	20	21	38	0
4	707	2	5	0	11	20	21	38	39
6	707	1	4	0	13	19	21	0	0
2	707	2	5	0	11	19	21	35	0
1	707	1	4	0	13	19	21	0	0
7	707	2	5	0	13	19	21	0	39
34	707	1	4	0	13	19	21	0	0
16	707	3	6	0	11	20	21	38	0
1	707	1	4	0	11	17	21	0	0
3	707	3	6	0	13	19	21	0	0

Figure 4.12. Example of sequential behaviors with different duration

It can be observed that the coded person (Participant No. 707, female) displays no agreement or disagreement behavior for about 98 seconds (code "0" in the "Agreement/disagreement" column), keeps her head straight for about 28 seconds (code "11" in the "Head pos." column), then puts her head on the hand for another 6 seconds (code "13" in the "Head pos." column), then shifts again to a straight position, then again on the hand for 42 seconds, and again straight. Meanwhile, the position of hands alternates in a different rhythm (column "Hand nand arm pos."). The body posture stays unchanged, oriented towards the speaker (code "21" in "Body pos." column), whereas the person performs several complex movements (in the "Body movements" column) that also occur in different combinations with all previous actions. The last column shows that the person also alternates moments of silence with conversations.

In this example, two behaviors stay completely unchanged and flow uninterrupted while the person changes her gaze direction, moves and converses repeatedly: (1) the person shows a rather neutral attitude (no agreement or disagreement behavior) and (2) she keeps the same body posture. Table 4.7 summarizes all changes in her nonverbal behavior display compared to the whole corpus.

Table 4.7

Comparison between changes in nonverbal behavior display of person 707 and the whole corpus

Behavior category	Change frequency for person 707 (instances)	Change in whole corpus (instances)
gaze direction	15	897
agreement/disagreement	0	14
head positioning	6	48
hand & arm position	7	318
body posture	0	28
complex body movements	10	349
conversations	8	202

The changes calculated for this person mirror in great measure the ones recorded for all behaviors in the whole coded corpus. It turns out that in the coded 10 minutes, people had a very unsteady gaze direction, whereas they rarely changed their head or torso postures.

Another perspective on sequential behaviors is to look in particular at gaze behaviors. There are five main gaze behaviors: looking to the front, looking down, looking to the left or right, staring into space, or keeping the eyes closed. Most of occurring patterns included looking to the front, down, and to left or right. These gaze directions alternate with each new coded line. Interesting patterns occurred, when people repeat their gaze directions for longer periods of time, as shown in Figure 4.13 depicting a student taking notes.

The sequence in this example starts with the student looking at the lecturer at minute 18:27. Two seconds later, the student starts taking

Minute 18:27   Minute 18:29   Minute 18:37   Minute 18:39   Minute 18:47
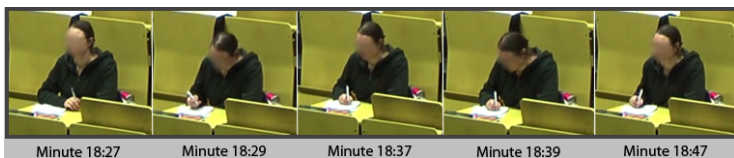
Figure 4.13. Example of sequential behavior

notes, then looks up again and down again. In this particular examples, the duration of note taking was longer than that of looking at the speaker. In the virtual audience case, such animations can have various durations and note taking can also be animated to last less than looking at the speaker, if this serves the training purposes.

The main role of sequential behavior analysis was to provide some information on how behaviors of virtual humans can be ordered so that they look natural. It is clear that people employ various nonverbal modalities to express attentiveness and inattentiveness. However, behaviors would have to flow one into the other. This progressive movements fluidity has been described in literature (Lourens et al., 2010) and in the design of embodied conversational agents (Hartmann, Mancini, & Pelachaud, 2006), but not specifically for attentiveness and inattentiveness. Furthermore, the analysis showed that nonverbal behaviors are not block-like. Modalities have different durations and these differences should be considered in the animation process as well. The findings here show how animations could flow from one behavior combination into another, by changing only the behaviors that are more prone to shift and leaving the other ones unchanged.

The calculation of how frequently behavior change can also help assess the effort put into design. Furthermore, frequencies can also serve future research into audience behaviors and also help conceptualize audience activity and passivity based on how often people change their gaze direction, postures, etc.

4.4.4.3 Duration

The coding was performed against a timeline, which permitted the calculation of behavior combinations' duration. The aim was to establish a baseline to time the animations in the design phase. Durations were recorded for each combination and means were calculated for combinations and not per person. Also, the mean was calculated for all behaviors, independently of their occurrence frequency, across the whole corpus of 966 behaviors. The mean duration of coded behavior combinations was $M = 7,9$ seconds ($SD = 15$ seconds). This high standard deviation can be explained first by the fact that several behaviors were coded for a very long time, with people keeping the same gaze direction for minutes. Apart from that, more than a quarter of the coded behavior combinations were only a second long, due to the fast shifts in gaze direction. Just like unique behaviors proved to be very hetero-

geneous with most of them occurring only once, their duration is also very heterogeneous. Moreover, more than two thirds of the behaviors last less than 7,9 seconds which represents the mean of durations.

Note taking was one of the activities that depicted attentiveness and lasted on average $M = 13$ seconds ($SD = 12$ seconds), with the longest duration of 48 seconds and shortest one of 11 seconds. Attentiveness which manifested through direct gaze to the speaker lasted on average $M = 6$ seconds ($SD = 13$ seconds), between one and a maximum of 110 seconds.

To calculate duration of inattentiveness, all behaviors that had eye gaze away from the speaker, independently of action trigger were grouped together. Calculations show that a person was on average inattentive for $M = 8$ seconds ($SD = 17$ seconds). This number is between the range of one second and 176 seconds. The latter represents the longest duration of someone typing on a laptop continuously.

It is interesting to note that coded persons spent less time looking at the speaker (six seconds) than looking elsewhere (eight seconds), and took notes for far longer periods of time (13 seconds) than both previous behaviors. The main advantage of having calculated behavior durations is that these help diversify behavior designs. Not only can virtual humans be assigned any of the almost 1000 coded behavior combinations, but each combination can have a different duration, which might have a different effect on the speaker. This variation can help design future studies and test hypotheses on whether behaviors with different durations have different effects on the listeners. Moreover, this can be used by trainees as well, if they will wish to customize their virtual audiences to display particularly long or short behaviors, in a realistic and empirically-grounded way.

Parallel behavior combinations coded here borrowed the length of the eye gaze, because eye gaze was the discriminant element for starting coding a new behavior combination. In reality though, each behavior can have a different length. For instance, a person can order objects while looking in various directions. Whereas the gaze changes, the action of ordering objects is the same each time the person shifts their gaze. Following the coding rule that eye gaze and/or head behaviors determine the duration of the coded combination all behavior combinations are given the duration of the eye gaze and/or head behavior. This can lead to behavior combinations where certain

actions are unnaturally segmented by eye gaze direction shift. The sequential behavior analysis proposed a solution to avoid this artificial segmentation of behaviors and regard them in a contextual manner.

4.4.4.4   Coding saturation area

The number of identified unique combinations (225) covers only a fraction of all possible unique ways that people in the recorded audience can manifest, based on the given codebook. The codebook included a total of 43 codes for behaviors, divided into eight categories. If all possible combinations are calculated, the eight categories with 43 total variables could give approximately 75.000 possible unique behavior combinations. In order to ensure that a small chosen sample is satisfactory, a few calculations were further performed to see whether the chosen sample can provide diverse behaviors.

Dividing the 225 unique combination to 14 coded persons means that, on average, each person expresses their attentiveness and inattentiveness in 16 different ways. These combinations are not equally distributed between people and vary depending on how active or static the coded person was. An active person can move the torso, the head, can type on the mobile phone, pass on objects, sneeze, scratch, leave the room, or engage in conversations. A less active one shifts eye gaze less often, and receives few codes for movements, hence the number of possible *new* behavior combinations is lower.

Taking these differences into account, a calculation was conducted to see when behaviors start repeating and whether adding new people to the analysis enriches the pool of unique behaviors and what is the rate of this process. The aim was to identify the point where one reaches a "saturation area" of new behaviors (where every new person adds only little to the whole pool of unique behaviors). This was defined as an area where the slope takes a less steeper angle, until, ideally, it reaches a parallel line to the X axis. Ideally, this means that each new person brings less new unique behavior combinations to the corpus until when no other new behaviors can be added, even if a different person is being coded.

To check how the coding of every new person enriches the pool of coded behaviors, all identified combinations for the 14 coded students were run through a MATLAB script. Figure 4.14 shows that the number of new unique behavior decreases almost linearly, with each
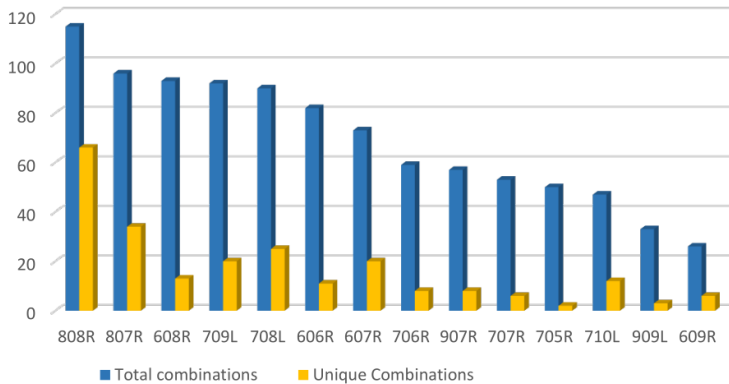
Figure 4.14. Calculations of saturation area from most active (808R) to least active (609R) coded person

newly added person. The first person (Person 808R) is the most active person in the coded corpus (115 total behavior combinations). The more active the persons, the more unique behavior combinations they will bring. Out of these 115, Person 808R displays around 60 unique behavior combinations. Conversely, the least active a person (static), for instance Person 609R, the least unique combinations they will bring. An interpretation is that, very active people vary their gaze direction and with it they also perform various body movements (see also section on activation). The more categories are coded for a person, the higher the chances that new behavior combinations will appear compared to less active persons who may not move their bodies at all. Hence least active persons, who are at the tail of the graph bring least new unique combinations.

It must be considered though, that the probability of behavior occurrence depends on the type of coded audience and coding instrument. Current findings about the saturation area can also be a result of the codebook, which contains many variables for body movements as well as for inattentive gaze direction. A shorter guideline could help reach the plateau faster, but at the cost of behavior diversity. Also, the algorithm for identifying the place when the slope starts decreasing begins with persons who are very active and moves on to less and less active people. If the algorithm uses a different selection method (e.g., random selection of first coded person), the plateau area might be somewhere

else or the slope might look totally different. The saturation area was calculated also to help establish how could coding be made more efficient, in order to grasp behavior variety. If the coding starts with people who appear very active, many of the diverse behaviors will also show up at the beginning of the analysis. Ideally, as coding moves to less and less active people, the number of new behavior combinations drops almost constantly.

CHAPTER 4.5

Conclusion

The study reported here was a continuation of the previous research which addressed specifically the potential characteristics of the virtual humans found in the first layer of the audience research model. Expert interviews revealed how important were demographic and behavior diversity. When asked what kind of audiences they've experienced during public speaking training or competitions, experts recalled pretty heterogeneous listener responses. These were grouped in the previous study into positive and negative ones and contained descriptive actions, like many studies on audience in VR (H. V. Anderson, 2003; Harris et al., 2002; Pertaub et al., 2002; Slater et al., 1999). These present a virtual listener's action in general terms: conversing, looking bored, applauding, yawning, nodding, talking to each other, verbalizing discontent, or being supportive by smiling or nodding. The observational study conducted here looked at audience behavior in detailed and answered two research questions: the first question regarded the way attentiveness and inattentiveness looked like within a student audience during a lecture; the second question looked at what behaviors within the audience triggered reactions of fellow students, as group, and how such behaviors spread.

Several nonverbal behaviors were retrieved from literature and helped build a basis for the analysis. These, together with others identified in a pretest with a different student audience led to a codebook that included postures and movements of the upper body (head, torso, and hands). Three main behaviors were sought after: attentiveness, inattentiveness, and group dynamic. Keeping an audience interested and attentive represents one of the key skills in public speaking (Verderber et al., 2010). Therefore, it is sensible to conceive a virtual audience that helps speakers achieve such a skill. Attentiveness and inattentiveness were differentiated on a single category–eye gaze/head direction, and every other simultaneous behavior that was associated with a certain gaze direction was assigned to either the attentive or

inattentive list. Dynamic behaviors were also observed and these included triggers that involve more members of the audience, such as knocking on the desk at the end of the lecture or involuntary attention towards a person leaving the room.

For the analysis, a student audience of 37 students was observed during a 40-minutes lecture at Technische Universität Ilmenau. The final analysis was made based on a sample of 14 students who produced a total of 966 behavior combinations. Findings showed that these can be reduced to only 225 unique behaviors (behaviors that differ from each other completely), which means that on average, every other forth coded line occurs somewhere else in the corpus (either for a different person or for the same person, but at a different moment in time). Out of all 255 unique behaviors, 119 occur only once in the coded time frame, which hints towards high behavior heterogeneity. This can be seen as an advantage that meets the needs of experts, by providing a diverse pool of manifest nonverbal attentive and inattentive behaviors.

4.5.1 Summary of findings

4.5.1.1 Attentiveness, inattentiveness, and group dynamic

Attentiveness and inattentiveness manifest in many various ways, with people performing a lot of movements and being far from completely still. This happened even if it was expected to see little activation from attentive people compared to inattentive ones. Not only that people performed various movements while looking to the speaker and paying attention to what was happening in the front of the lecture hall, but their behaviors were, as seen, highly heterogeneous.

Attentiveness involved gaze and body orientation towards the lecturer and sometimes also taking notes. Inattentiveness on the other hand manifested through performing various activities that relocate the attention resources away from the speaker, such as using the mobile phone or engaging in conversations with other students. Further behaviors were avoidant gazes and shifts in body postures that were accompanied by fidgety movements. Out of all coded behaviors, most behaviors belong to the inattentiveness category.

In the current study, both attentiveness and inattentiveness occurred in a highly varied fashion. The factor that helped differentiate between

these two behaviors in the coding process and analysis was gaze and/or head movement. Moreover, additional things can be said about people who payed attention or not:

1. Activation levels: Both attentive and inattentive people performed body movements. They ate, drank, scratched or groomed themselves, ordered the objects on their desk and did all this while looking either directly to speaker or ignoring him completely (Figure 4.15). However, people tended to move less when they were looking to the speaker and to move more when they were looking elsewhere.

2. Duration: Attentiveness lasts on average six seconds, while inattentiveness eight seconds. The difference is not large, but throughout the coded corpus, duration of behaviors was highly heterogeneous, from one second to 176 seconds. The longest lasting attentive behavior was "taking notes", which lasted on average 13 seconds. Apart from this, most of the coded combinations lasted only one second. The longest lasting inattentive behavior was coded when a person kept looking down to the desk for about three minutes continuously.

Another finding described how group behaviors occur and pointed out that people can distract others and make them shift their attention away from the speaker. When people come late or leave the lecture hall, as they walk along the aisle, they gradually enter the visual fields
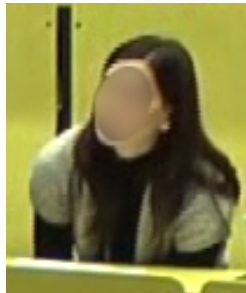
Figure 4.15. Example of a very active person who looks at the speaker while conversing with a neighbor and unpacking her bag

of audience members. As these turn their heads or become distracted, the behavior seems to engulf the whole group.

Another behavior that was observed to start in low numbers and then spread throughout the audience was knocking on desk at the end of the lecture, which didn't occur simultaneously, but spread in a few seconds across the whole lecture hall.

4.5.1.2   Parallel and sequential behaviors

All attentive, inattentive, and group dynamic behaviors can be structured into unique behaviors (appearing only once), behavior patterns (appearing at least two times), and parallel and sequential behaviors. Behaviors patterns reveal that some actions were performed the same way over and over again, such as the case of the most frequent sign of attentiveness which appeared for 72 times. In this corpus, such a behavior can be considered representative and safely used in the animation corpus for virtual humans.

Parallel behaviors revealed that most ones included a balanced mix between static and active behaviors (six categories per behavior combination). Extremely active people who were coded for all eight possible behavior categories as well as those very inactive ones were rather rare. Almost half of all coded behaviors contained some body movements.

Next, the analysis looked at how behavior combinations follow one another, in order to establish a possible order of behavior animation. Human nonverbal behaviors follow a certain flow and movements change into other movements at different speeds (Pelachaud, 2009). In order to transfer this quality of movement fluidity to the animation of virtual humans, the analysis looked at how behaviors followed one another on the time line for the identified nonverbal categories. Eye gaze change rate was the most frequent change. Some of the behaviors that rarely changed were body posture and head position. Most people sat facing the speaker and kept their head straight. These findings are important for the animation endeavor because they help establish a hierarchy of behavior frequency. Such a hierarchy can help simplify modular animations. Findings showed that people kept their body posture orientation rather constant during a lecture and this feature can be kept constant during animation processes of virtual humans. The other ones that vary more often in real life (such as eye gaze/head

direction or body movements) can also be varied at similar rates for virtual humans.

4.5.2 Discussion

The purpose of this observation study was to provide a pool of diverse nonverbal behaviors that depict attentiveness and inattentiveness towards the speaker as well as group dynamic, in a presentation situation (Audience level in the research concept). By this, it addressed the first level of the audience concept model and enhanced the pool of virtual humans characteristics with empirically-grounded behaviors.

Compared with a similar study conducted by Kang et al. (2013) where postures were coded, implemented within an audience, and then evaluated by users, the approach here was to identify from the beginning attentive and inattentive behaviors within an audience, based on gaze direction cues as discussed in literature (Frischen et al., 2007; Norton & Pettegrew, 1979).

The analyzed sample was rather small for a content analysis both in terms of coded person and in terms of time segments, but it's comparable to the sample used by other researchers who observed audiences. Kang and colleagues (2013) coded the behaviors of 16 listeners for the duration of four different presentation types that lasted no longer than seven minutes each. They report a behavior corpus that contains around 300 unique behaviors and their final analysis is conducted on 59 postures (compared to 225 unique behaviors that were identified and also analyzed here). The data corpus proved rich enough to allow performing several further analyses, such as to look at how complex displayed behaviors appear to be. Even if all actions that people displayed could be grouped into attentiveness or inattentiveness, not all behaviors within these groups were equally complex, some contained more modalities than others, some contained more movements than others. This finding is relevant for the animation procedure, because it can help simplify and prioritize the absolute necessary behaviors that have to be animated, while still keeping the meaning of the displayed action. For instance, inattentiveness doesn't have to be highly complex, it can only mean that a person looks elsewhere, while sitting completely still. An attentive speaker can also be very fidgety and engage with many other activities, while

simultaneously looking to the speaker and giving the impression that they are listening carefully. The main finding of the complexity analysis was a spectrum of attentive and inattentive behaviors from most simple to most complex ones. This should help prioritize the behaviors that have to be animated, as well as help decide on their necessary complexity before the begin of the animation process. Negative behaviors can be as simple as positive ones, as long as the discriminating elements are visible and clear.

Whereas attentiveness was defined as a clear gaze orientation towards the speaker, inattentiveness was conceptualized from two perspectives, depending on the action trigger. A person could be deliberately inattentive or could be distracted by someone else. The trigger was coded here for each person, to establish whether inattention was deliberate or not. The findings showed that people were mainly choosing deliberately not to pay attention, than get distracted by someone else from the audience. In the absence of this deliberate choice and of disturbing factors, people would return to pay attention to the speaker. An autonomous virtual audience, as proposed by Kasap and Magnenat-Thalmann (2007) would mirror these three triggers and would be able to respond both to other virtual humans and to the speaker. It remains to be see though where the threshold of attention lies and what decision would a virtual human take, when different triggers (from speaker and other virtual humans) occur simultaneously.

Related to involuntary attention are also group behaviors that appear when triggers affect several persons (Farkas et al., 2002; Hylton, 1971). Group behaviors, such as an audience that gradually changes from quiet to restless, have not been reported in the literature on virtual audiences in phobia treatment and the study here offered two examples of how such behaviors look like. Due to the scattered locations of the students who were seated in small groups of two-four people, it was difficult to observe behaviors that would spread throughout the whole audience, since people were separated by empty rows. However, they gradually got distracted when colleagues entered the room and moved along the aisles. A denser group could have reflected group behaviors in a better fashion.

Managing audience attitude is a key training factors in public speaking coaching, as it prepares speakers for unpredictable listeners whose attention they have to grab and keep (Verderber et al., 2010). The

behaviors identified here could mirror genuine interest or disinterest attitudes, but such attitudes can only be uncovered by asking students how they actually felt or by actually measuring attention and inattention with standardized memory or knowledge tests. The findings here do not wish to equate observed behaviors with felt attitudes, but simply show what people do when they sit and listen to someone giving a presentation.

4.5.3   Limitations and outlook

A first limitation of the study lies in the very specific observed audience. This was a student audience filmed during a lecture at the university. A different population, in a different location could have led to other findings.   The location in particular imposed various limitations on observed behaviors. The lecture hall had the shape of an amphitheater and only the upper part of the torso of listeners was visible. Legs and feet could therefore not be coded.   Additionally, a speech other than a lecture could have triggered other behaviors. Kang and colleagues (2013) used a positive, a neutral, a boring, and a critical presentation as independent variables and measured differences in audience attitudes that reflected the valence of the presentations.

A methodological limitation represents the coding procedure and rules for the observed behaviors. Previous studies used a fix time frame and coded behaviors that occurred within this frame (e.g., as SKANS V, Gifford (1994a)). In the present study, the coding was dynamic, based on the rate of gaze direction change. The study didn't count the frequencies of each behavior from each category, but the number of behavior combinations. This different of focus led to the creation of a customized coding platform that can record behavior combinations on a timeline. Furthermore, coding behavior combinations proved strenuous and required replaying video segments for several times in order to identify the exact combination. All in all, the dynamic coding solution is laborious and requires a good training, but it has a major advantage over other coding methods, because it captures behavior simultaneity as it occurs in the natural setting.

A limitation that can be corrected in future studies is the lack of sound coding. The camera recorded sounds as well, such as various ambient sounds (e.g., sounds for shuffling objects, for chairs lifting as students

stood up, etc.) and one can hear students whispering. Except for knocking on the desks at the end of the lecture, no sounds were coded, because they were too faint and unclear to permit an unambiguous coding and correlation with each coded person. For the virtual audience, such sounds can be either coded in future video recordings or can be sampled separately and matched with the behaviors. As well as movements that trigger involuntary attention, sounds can have the same effect on a person's attention resources (Eimer et al., 1996), therefore sound (e.g., laughter, opening or closing doors, phone ringing) could be used in the virtual audience to trigger reactions among the virtual humans.

The coded behavior combinations can be used in the animations as they have been described here, but could as well be further adapted, with regard to their intensity. Future studies could measure how intense such behaviors appear to people and whether one can speak of *degrees* of attention, based on how intense certain movements are, instead of whether a behavior appears or not. Since intensity of each behavior or each person was not measured here, a way to solve this aspect for the virtual audience would be to regard intensity at group level: vary the percentages of attentive and inattentive people in the audience and measure whether speakers perceive different degrees of audience general attentiveness based on how many people pay attention or not. This calls also for the development of a measurement instrument to identify discriminatory behaviors that make audiences appear attentive or inattentive.

Apart from the technicalities of coding that can be improved in the future, the categorization of observed behaviors in the observed audience could also be backed by participant accounts on how they actually felt during the recording time. At first glance, the majority of the people were inattentive, but did they actually feel this way? Knowing the way people actually felt during the observation would reduce the risk of over-interpretation of observed behaviors. Moreover, knowing how people feel could also help differentiate between idiosyncratic behaviors (Ekman & Friesen, 1969) and behaviors displayed by everyone.

Since the used sample was small and contained also students, results are difficult to be generalized to other audience types. A different location, audience membership or even a different organizational setting (the same group of students in a seminar) could yield different patterns

of attentiveness and inattentiveness. Even if future research analyzes a different setting, it is expected that there are similarities between those findings and the ones provided here. Eye gaze direction should have the same values like here (towards the speaker or averted); conversations should also occur in similar fashion, as well as body postures (e.g., to the front or averted). Most differences could appear in the category of complex body movements. Movement types presented here were specific to students (based on literature on student attentiveness and on the pretest with another student audience). These could vary with the setting and with manipulations of the objects that are around the coded person. In a room without desks, object manipulation could look totally differently and people might have no place to order their things and simply lay them on the ground or on nearby chairs, as it often happens at conferences. Nevertheless, in spite of the limitations mentioned above, the results of the current study can serve as a starting point in the creation of a large behavior database for audiences in virtual reality training settings. Last, after the virtual audience is endowed with nonverbal behaviors as proposed here, speakers could also evaluate what behaviors they actually notice while giving a speech. This would help improve design by selecting the most visible behaviors, as well as the most relevant ones, from the perspective of speakers.

5

Virtual audience design concept

CHAPTER 5.1

Introduction to the design concept

The present chapter synthesizes the main findings of all studies presented before and introduces a design concept for a virtual audience. The concept addresses the layers of the audience research model and contains descriptions of virtual audience characteristics, their behaviors (illustrated by an actual scenario with attentive and inattentive behaviors), and a list of proposed features for a graphic user interface.

The literature on virtual audiences used in phobia treatment reports information about the simulated characters, but rarely explains in detail how scenarios were created, such as physical features of virtual humans, or the virtual rooms they populate. The major feature that is usually being discussed is behavior valence (positive or negative), which is used to trigger and help measure anxiety in people with phobia (P. L. Anderson et al., 2005; Pertaub et al., 2002). As seen though in the communication literature (Verderber et al., 2008) and in the discussions with training experts, audiences have great potential for customization beyond the usual gender ratio and positive/negative behavior valence. Audiences have history, they belong to a culture, carry expectations, beliefs, likes, and dislikes. They can be familiar to speakers or not, they can be equals or occupy different hierarchical positions in their jobs. Furthermore, audience members can come from the same discipline as the speaker or from a different one, and speakers are often required to learn to address people outside their audiences, such as potential clients or investors (Dannels & Housley Gaffney, 2009).

Another element that ought to be brought into the discussion is the graphic user interface. Therapist control is a pervasive topic in many studies on anxiety treatment in VR (P. L. Anderson et al., 2005; Wallach et al., 2009; Wiederhold & Wiederhold, 2005). However, little is known about the full potential of the application and its GUI. A quick review reveals that the most used option is audience behavior

selection (i.e., selecting how audiences behave at a certain point in time) (P. L. Anderson et al., 2005; Wallach et al., 2009; Pertaub et al., 2002). Nevertheless, recent studies have started to focus on applications that are easy to use and to implement, such as the NeuroVR2 used in health care which permits virtual scenario customization for 14-pre designed virtual environments (Riva et al., 2011). Other studies suggest new customization options for the audiences, such as adjusting scenarios for different performance purposes (Kang et al., 2013). Hence, there is rich development potential for an interface that can accommodate the needs of trainers.

In spite of the fact that the user interface was not specifically researched in any of the two studies reported in the dissertation, the interviews provided sufficient information to be able to synthesize a few useful interface characteristics.

This chapter summarizes all findings and starts with a theoretical and methodological part on pedagogical virtual environments design. This is followed by a section that describes the audience concept and a proposed audience scenario.

CHAPTER 5.2

Theoretical background on virtual environments design

There are two ways to carry out information technology design tasks, by following the technology-centered approach or the user-centered approach (Endsley & Jones, 2011). Whereas in the first technology lies at the center and and human users have to adapt using it in a way that may not be optimally suited for the task, in user-centered design (UCD) technology is molded to fit the requirements of the users and to help them fulfill their goals, tasks, and needs (Endsley & Jones, 2011).

Although researchers seem to have disparate opinions about the definition of UCD (Gulliksen et al., 2003), there is agreement on its guiding principles. One of the principle is to organize the technology around the goals, tasks, and needs of the user and to consider these early in the design process (Endsley & Jones, 2011; Gulliksen et al., 2003). Gulliksen et al. (2003) suggest even that designers should get to know real or potential users, and that users should become active in the development process. Another principle is to base design on the way users process information and make decisions in the real world so that they can use the same mental schemes in solving tasks with the help of technology (Endsley & Jones, 2011). This is further complemented by the control and awareness principle, which ensures that the system helps the user stay aware and in control of the situation where they perform (Endsley & Jones, 2011). Other principles refer more to the design methods and processes, such as simplicity, prototyping, step-wise dedicated design activities, as well as incremental and iterative design (Gulliksen et al., 2003). The next sections will present the theoretical background and methods in use for virtual environments design. These help apply the UCD principles to the findings of the empirical studies in order to transfer them to the audience design concept.

5.2.1 Technical affordances

The design concept is constructed with the purpose of serving as train-
ing material in virtual reality for public speaking training. Its role is
to offer both trainers and speakers a tool that facilitates and diversi-
fies speech practice sessions. Virtual environments used in education
are called Educational Virtual Environments (EVEs) or Virtual Learn-
ing Environments (VLEs) and are defined as environments "based on a
certain pedagogical model, (that) incorporate or imply one or more di-
dactic objectives, provide users with experiences they would otherwise
not be able to experience in the physical world, and redound specific
learning outcomes" Mikropoulos and Natsis (2011, p. 770). These can
take the shape of desktop as well as immersive environments such as
head-mounted displays and CAVEs. Mikropoulos and Natsis (2011)
reviewed the literature on EVEs published between 1999-2009 and dis-
covered that, compared with social sciences where abstract ideas are
hard to represent in VR, many studies on VR educational applications
are in natural sciences and mathematics. Another finding was that the
studies focus rather on the technological approaches of VR and not on
the pedagogical ones. They form a list of features belonging to VR
that have been used in literature and have been found to influence the
effective learning outcome. These are "first-order experiences, natu-
ral semantics, size, transduction, reification, autonomy, and presence"
Mikropoulos and Natsis (2011, p. 770). Although gathered across var-
ious teaching domains, some of these characteristics can be identified
to match the concept proposed in this chapter. There are some excep-
tions however, such as transduction, reification, and autonomy. Trans-
duction represents the process of translating information that cannot
usually be accessed through senses into accessible elements, such as
showing the world through infrared light otherwise unavailable to the
human eye (Winn, 1993). Because transduction is presented as a sen-
sory translation of existing visual and aural input, it applies rather to
the natural sciences than to human communication experience, such
as public speaking. Further on, through reification, something that is
abstract and out of the human experience, such as chemical processes
in the body can be represented in a comprehensible way through inter-
active representations in VR (Alverson, Saiki, Jr., Caudell, Goldsmith,
& Stevens, 2006). Communication is within human understanding and
experience, therefore the feature is not applicable to the audience con-
cept here. Autonomy is represented as a feature of shared virtual envi-

ronments that have to function also when only a part of the users are present (Mikropoulos & Natsis, 2011). The proposed virtual audience is intended to function for each speaker at a time, therefore autonomy doesn't currently apply. The other four remaining features will be discussed with regard to their compatibility with the virtual audience concept.

First-order or first person experience represents the actual speaking practice in front of the virtual audience that each trainee would have to undergo. Through natural semantics, information is displayed in a form that can be easily recognized (Bricken, 1989). This feature manifests through the actual virtual audience, which does no longer have to be imagined (e.g., image speaking in front of a restless audience) or simulated through role play among trainees (e.g., fellow trainees who play a restless audience), but can actually depict realistic diversity of audiences, through looks and behaviors. Changing the size of the visualized object or accessing out of ordinary scaled objects in a manageable size in VR, such as galaxies or atoms can as well apply to communication training–more precisely to the virtual rooms hosting the listeners. The audience concepts includes three room types that differ by size, as well as by seat arrangement. The last element is presence and some of the factors that were found to influence presence in learning virtual environments were sound (Mikropoulos & Strouboulis, 2004; Whitelock, Romano, Jelfs, & Brna, 2000) and navigation with many degrees of freedom (in a study with pupils, by Mikropoulos and Strouboulis (2004)). In relation to the virtual audience concept, presence is conceptualized as a result of a user's interaction with the virtual audience and is expected to occur once the virtual audience prototype is put into practice. It can be then concluded that the literature on VLEs contains features that have as well emerged through the empirical studies conducted in the dissertation and have been integrated in the concept proposal here, bringing it in line with the current state of the art.

Dalgarno and Lee (2010) propose a theoretical model of learning in 3D virtual learning environments by highlighting the two unique characteristics of 3D virtual environments that enforce their pedagogical potential: representational fidelity and learner interaction. The first one refers to visual and aural representation of the content, as well as to the consistent behaviors of the displayed objects which integrate well into their virtual environments (similar to the autonomous behaviors

proposed by Kasap and Magnenat-Thalmann (2007)). Learner inter-
action refers specifically to the verbal and nonverbal exchange between
learner and virtual environment, as well as to navigation, object manip-
ulation and construction, and environment attributes' and behaviors'
control. The researchers argue that these attributes would help learners
construct their identities in the virtual environment, experience pres-
ence, and enjoy co-presence with fellow learners and ultimately achieve
learning benefits through a series of processes, such as contextual learn-
ing, collaborative learning, or engagement. Dalgarno and Lee (2010)
suggest that their model can serve as research agenda for the design
and use of learning virtual environments. The representational fidelity
and learner interaction are already two pillars of the audience concept
discussed here that match the first level of the research model (virtual
audience characteristics and behaviors) and the second level (speaker-
audience interaction). The processes and theories that help explain how
users may acquire communication skills through the audience concept
are presented in the following section.

5.2.2 Skills transfer

The studies reviewed above focused mainly on general visual and aural
characteristics that foster learning in virtual environments. A more
focused approach comes from the literature on communication skill ac-
quisition, which proposes its own models of skill teaching. The medical
field is rich in doctor-patient communication studies and Brown and
Bylund (2008) point out that the literature on skills in the field is am-
biguous with regard to what abilities have to be taught. Therefore, they
suggest a comprehensive model that helps standardize skill teaching and
assessment. According to the model, the training has to consider sev-
eral elements, such as communication goals (the desired outcome of the
communication) and strategies (a priori plan of steering the communi-
cation towards the goal), communication skills (understood here as unit
of speech, such as summarizing information), process tasks (speaker be-
haviors that create an effective communication environment), cognitive
appraisals (interpretation of verbal and nonverbal listener feedback),
patient cues (indirect cues given by listeners about their immediate
information needs), and patient barriers (such as undisclosed patient
perceptions, usually fear-induced). This reminds of the microteach-
ing approach where social skills are broken down into components that

can be addressed separately (Hargie, 2006). A more general division is skill preparation, training, and evaluation (Hargie, 2006). Both the detailed skills approach by Brown and Bylund (2008) and the components of Hargie (2006) are reflected in the audience concept proposed here. First, through the user interface, trainers together with learners can establish the communication goals, strategies, and skills that have to be acquired. Second, after taking these steps, with the help of the highly customizable audience in terms of characteristics and behaviors and of the audience scenarios setup, learners can practice process tasks, cognitive appraisals, identify listener cues, and deal with any other audience communication barriers that could emerge in a public speaking situation. Furthermore, in the broader context of public speaking training, the virtual audience and scenarios come into play after the skill preparation phase (e.g., theory) and before evaluation (e.g., through standardized tests).

When it comes to virtual environments that function as learning environments, it is necessary that the environment design follows learning theories that help fulfill the educational goals. Mikropoulos and Natsis (2011) found in their literature review only a few studies mentioning theoretical approaches to learning and, when mentioned, the most common theory is constructivism. The constructivist view assumes that "knowledge is individually constructed and socially coconstructed by learners based on their interpretations of experiences in the world" Jonassen (1998, p. 217). Constructivist learning environments reproduce in the virtual world specific activities and tasks for which learners apply the same skills they would apply in a real world situation (Mikropoulos & Natsis, 2011). But constructivism doesn't only apply to object manipulation, virtual world navigation, and conceptual understanding of abstract worlds, like in the review of Mikropoulos and Natsis (2011). It is a theory related to communication as well, and from this angle it supports the audience concept design given here. In communication, constructivism seeks "to explain individual differences in people's ability to communication skillfully in social situations" Delia (2003, p. 98). According to this theory, people have different communication competences based on their own cognitive complexity (Delia, 2003). Moreover, Delia suggests that people who are competent communicators tailor their messages based on their awareness and perceptions of the recipients. He calls this "person-centered messages" (p. 101). The successful message rests on a three-step approach to crafting

the communication: choosing the goals–what is the goal of the communication, selecting the plans–apply a procedure of steps to follow to fulfill the goal, and acting–delivering the message. These steps resemble the ones mentioned earlier in skills training for medical personnel and can as well be identified in the communication effectiveness model given by Verderber et al. (2008). It appears legitimate then to chose constructivism as the representative theory for devising a highly customizable virtual audience for communication training purposes. Through its main core of audience customization and interaction, it ought to fulfill the learning needs of trainees and to help achieve these goals through a theory-based design.

The chapter here presented a few theoretical approaches of virtual learning environments that can serve the design endeavor. Communication skills are inherent to human experience, but they can be improved through constructivist approaches. The design of learning environments has to consider not only the technical elements, but also the learning theories that maximize the knowledge transfer potential. Moreover, a user interface that supports human-computer interaction should reflect user-centered principles. The next section will offer a methodological approach for turning these theoretical frames into actual design steps and final product suggestions.

CHAPTER 5.3

Methodology

5.3.1 Design models

Virtual environments have been applied for various educational purposes, with diverse users, learning contexts and pedagogical targets (Sánchez et al., 2000). However, it's agreed in the literature that there is little consensus altogether on virtual environments specifications, design processes, and testing (Eastgate, Wilson, & D'Cruz, 2014), and scientists call for unifying guiding models (Sánchez et al., 2000). Moreover, there is a lack standardization discussed by Riva et al. (2011, p. 494) who conclude that the limitations of VR in psychotherapy and behavioral neuroscience are lack of standardized hardware and software with limited customization options, low availability of standardized protocols, high costs of design and testing VR prototypes, and little user-friendly VR applications.

To address this problems, Sánchez et al. (2000) suggest a metaphor specifically for VR and which ought to be universally applicable in education. The model is intended as a guideline for the design of virtual environments independently of their learning purpose. The authors argue that design should start from the real environment and source knowledge to be transmitted to students. These are then mapped into the virtual environment with the help of a "metaphorical projection" into four different planes. This metaphor ensures first that the knowledge is structured in a familiar way to users within the VR (structural plane), that the pedagogical approaches are represented as well, through didactic content and roles of teachers and students (learning plane). Next, the metaphor addresses navigation, and how users move in the virtual environment, as well as the student's viewpoint (navigation plane). Following, the last element is interaction and the way users interact with the virtual scenario, the depicted objects, as well as with other users (interaction plane). Learning

activities are planned to take place in the virtual environment, and the authors enumerate several representative features that facilitate this, such as navigation, scaling, changing viewpoint, networked learning, as well as psychological experiences of presence.

Another element of the model are the actors involved in the design process. These include not only designers but also end users, such as teachers and students. These should both participate in the development phase and also customize the final version of the system, once this is done (Sánchez et al., 2000). The final choice over what the virtual system is going to contain, for what student profiles (e.g., age, background, etc.) as well as through which pedagogical program the knowledge is to be transmitted belongs to the learning module, which determines also what elements from the real world have to be simulated, as well as how the virtual reality can be applied.

The model of virtual reality systems for education proposed by Sánchez et al. (2000) is complex and offers a good overview of design elements, actors, and sources for the content. It can also serve as framework for research. For instance, the empirical studies conducted here match some of the elements proposed by the model. At the structural plane, a virtual audience scenario with seated virtual people waiting for the speech to begin represents a familiar speech situation that trainees may be acquainted with. At learning plane, the expert interviews helped understand the active role of the trainer in controlling the virtual audience. At the interaction plane, it's speaker-audience communication and feedback that sets in motion skill practice and acquisition. Audience customization based on speaker's needs, as identified in the first empirical study, as well as the type of skill they practice (e.g., answering questions) reflects the learning module.

The model however is a theoretical framework and contains no clear methodologies on how to approach design. A solution that directly addresses training programs comes from Helander (2006). He proposes a model that synthesizes the necessary steps of such a program: (1) defining training objectives; (2) deriving training content; (3) designing methods and training material; (4) obtaining the training program; and (5) evaluating training graduates. There are of course similarities between these two models, with both considering the training content and training methods and materials. The second model however bears a temporal nature, which makes it easy to position the audience concept

and scenarios proposed here within a design flow. Therefore, the audience concept and scenario presented here address the second and the third steps of this model as they contain concrete virtual human characteristics, behaviors, and suggestions on technical aspects.

Although the models presented above are more like guidelines for research preceding actual design and not actual methodological procedures, they are representative for the way design is approached. A research conducted by Kaur (1998) who interviewed virtual environment designers showed similar steps in the design process. Experts mentioned clarifying the requirements (e.g., interaction requirements), gathering reference material from the real world objects, structuring the graphical model and, if necessary its division among designers, building the actual objects and positioning them in VE, and further enhancing the environment with graphic elements (e.g., textures and lightning). It becomes clear then, that once the initial research on educational purposes and source knowledge is done, the next step is the actual content creation. The virtual audience and scenarios content stems from data collected through social science methods and requires structuring in a manner that follows virtual environments design customs and standards.

5.3.2   Content creation

The audience design concept and the audience scenario introduced here are, to the knowledge of the author, the first of their kind, with regard to their detailed nature. Moreover, all the information given here is founded on empirical work. Due to the multidisciplinary approach of the studies, the design concept had to present the data in a structured and simple mode.

A common practice in the field is to present virtual humans design propositions as system architectures (Rist et al., 2003). However, due to the data variety (findings on listener behaviors, on trainer expectations, on design risks, etc.) and level of detail (lists with demographic and behavioral characteristics), a different data presentation path was chosen. The method used to sum up the findings was to create lists of elements that have to be designed. This method rests on a common practice in game design–the game design document (GDD). GDD represents the blueprint of a game and contains usually

detailed descriptions of characters, levels, and of technical requirements (Mitchell, 2012). Here, only a simplified version has been used, because some usual features of GDD are not currently applicable to a virtual audience for public speaking training, such as game play style, levels, or maps to navigate the environment (Mitchell, 2012).

The lists presented here contain, apart from the descriptions of elements in the audience concept, a few lines about the advantages and disadvantages of the suggested content that mark its educational value in the training process, as well as technical opportunities and challenges that balance design from the VR technical perspective (as they have been discussed with experts).

The scenario presented in the second part of this chapter represents a list of behaviors that the virtual humans could embody during a speech training session. This term should not be confused with scenario or script used in other contexts, where they represent informal descriptions of what users do with particular technologies and what tasks they perform(Moulin & Brassard, 1996). The intention here was to create reality-based content for public speaking practice and not to create scenarios with how trainers and students would use this content. Moreover, the scenario reflects observed behaviors and not an idealization or fictional recreation of behavior manifestations. To find a parallel in the literature, the scenario resembles the *scene* component of storyboards, which represents "a non-decomposable learning activity which can be implemented in any way" (Jantke, Knauf, & Gonzalez, 2006, p. 3175). The scenario given here is one such instance of how a virtual audience can behave and it is presented as an indivisible training element.

CHAPTER 5.4

Results

The order of the listed elements follows the levels of the audience research model. The first described elements are virtual audience characteristics, followed by their behaviors, and the virtual spaces they could inhabit. Next, the list continues with a graphic user interface which corresponds to the interaction level of the model. Due to the detailed accounts of how audiences could look and behave, as well as of how they can be customized through a user interface, the concept here responds to the customization need argued by Riva et al. (2011). Moreover, it responds to the constructivist view of adaptable learning environments and of communication skill (Delia, 2003; Jonassen, 1998).

5.4.1 Audience concept

5.4.1.1 Virtual human characteristics

The virtual human features reported in this section correspond to the first level of the audience concept. The theoretical model that guided the research comprised technical and human-like attributes that virtual audience members ought to embody. All audience features that emerged from the interviews are reported below. These include familiarity between speaker and audience, which has been previously researched in relation to anxiety levels (Maclntyre & Thivierge, 1995), yet, to the knowledge of the author, not in public speaking studies in virtual reality. Other reported features are age, sex, occupation, place of origin, outfit, religious affiliation, political views, and hobbies.

1. Relationship to speaker:

   (a) Known or unknown - work colleagues, fellow pupils or students, friends, relatives, complete strangers. Experts trained work groups in companies and organized open seminars for pupils or students. Familiarity with training colleagues depended on each seminar. Sometimes people knew each other from the beginning (e.g., a work group), sometimes they got to know each other during the seminar.

   (b) Different or equal in hierarchy - employees at the same level, higher, or lower as the speaker. Experts discussed training people to present in front of colleagues, in front of managers, in front of examination boards (i.e., at public speaking competitions).

Main advantages: People who know each other can support and encourage each other during training; speaking to a unknown audience is a good training exercise for performers who prepare for public speaking competitions (RM_TE).

Main disadvantages: It requires imagination. Having a known audience and training for an unknown one requires speakers to imagine they address someone else.

Technology perspective: It's possible to import photos or videos of familiar people and mesh them on virtual humans, but it's very difficult to animate them and make them behave as the person they would depict. This is risky for immersive environments because a mismatch between a real person and their virtual counterparts can harm realism (DT_VRTechE).

Further suggestions: This feature can be expanded to accommodate familiarity of audiences with a topic, and with a presentation style, as in communication across the curriculum and communication within the disciplines (Dannels, 2001, 2002), in order to train different rhetoric styles and data organization methods specific to different disciplines.

2. Age. Experts mentioned an age range from 12 to 60 years old for the speakers they have trained. Proposed age segments:

   (a) 21 and under
   (b) 22 to 34
   (c) 35 to 44
   (d) 45 to 54
   (e) 55 to 64
   (f) 65 and over

   Main advantage: Varying audience age teaches ways to adapt content and address it properly (e.g., avoid using language that might seem inappropriate to elderly) (CS_TE).

   Main disadvantage: None were mentioned.

   Technology perspective: The feature is available and has been implemented already in speech anxiety treatment in VR (BH_VRTechE).

   Further suggestions: Implement scenarios with various age mixtures, e.g., old and young, only young, only old, etc.

3. Sex: male and female

   Main advantages: Help people overcome anxiety in front of the opposite sex. Teach how to deal with taboo topics in front of the opposite sex (e.g., a female speaker says something about nakedness to a male audience with a conservative religious background) (CS_TE)

   Main disadvantages: None were mentioned.

   Technology perspective: The feature is available has been implemented already in speech anxiety treatment in VR (BH_VRTechE)

   Further suggestions: -

4. Occupation:

   (a) student:

      i. high school pupils
      ii. university students

   (b) employee:

      i. management function
      ii. non-management function

Main advantages: Offer potential users customized audiences that match their occupation status or audiences from which they differ dramatically. Teach how to present succinctly in business meetings (RM_TE).

Main disadvantage: None were mentioned.

Technology perspective: The feature is available and has already been implemented in fear of public speaking virtual applications (BH_VRTechE).

Further suggestions: Depict various occupations through representative clothing; have the audience pose content and domain-related questions.

5. Place of origin:

   (a) same as the speaker
   (b) different than the speaker

Main advantage: Teach cultural diversity and modes of addressing (CS_TE). Training experts mentioned country of origin and continents. Among discussed nationalities were British, Cameroonian, German, Romanian, and Russian citizens they've trained. With regard to ethnic groups one expert mentioned Arabs and Gypsies.

Main disadvantage: Can create confusion for someone who is not used to speaking to foreigners (CS_TE).

Technical perspective: Cultural contexts have already been implemented in military applications for negotiation skills training, such as the Enhanced Learning Environment with Creative Technologies (ELECT-Bilat) presented by P. Kenny et al. (2007).

Further suggestions: Use place of origin as potential language barrier, with audiences who have basic, moderate, and advanced language skills in the language used to deliver the speech.

6. Outfit:

   (a) casual

   (b) event-based: e.g., formal for business meetings

   (c) subculture fashion: e.g. an expert mentioned Gothic fashion

Main advantages: Speakers get accustomed to audience variety (CS_TE).

Main disadvantages: None were mentioned.

Technical perspective: Outfit customization is available and has been implemented already in audience design in virtual reality phobia treatment (Grillon et al., 2006).

Further suggestions: See point 4 above.

7. Religious affiliation, political views, hobbies:

   (a) same as the speaker

   (b) different

Main advantages: Trainers consider these elements relevant to help customize the audience and teach speakers to adapt to audiences with diverse expectations. Trainers also wish that virtual humans with different backgrounds along these lines challenge the speakers with comments and questions (RM_TE).

Main disadvantage: None were mentioned.

Technical perspective: Outfit and virtual location matched a particular religion in previous applications (a full mosque during prayer time in Ulicny and Thalmann (2002)). However, the cited study focused on crowd animation and not on user-virtual human interaction.

Further suggestions: See point 4 above.

These were the main audience characteristics that emerged in the interviews and they confirm the importance of audience background,

history, preferences, etc. as discussed in the literature on audience characteristics in real public speaking settings (Verderber et al., 2008). The list can be further expanded and virtual characters can be endowed with motive, personalities, moods, emotions that would explicate their behaviors in a training scenario.

5.4.1.2   Virtual characters behaviors

Although it is impossible to completely separate audience behaviors from the interaction with the speaker (second level of the model), the actions of the virtual listeners are reported here as part of audience characteristics (first level of the audience model), since no actual influence of a speaker's behavior was measured on the observed and coded attentiveness and inattentiveness.

This section contains lists of behaviors displayed by students in the observation study, which are divided into attentive and inattentive behaviors. The lists contain snapshots of persons within the student audience and descriptions of the behaviors they perform. Behaviors depicted in Table 5.1 represent the first 10 attentive encountered actions. The full list of occurring behaviors (the complete code sheet) can be consulted on demand.

Table 5.1

Attentive nonverbal behavior combinations - first 10 in descending order of frequency
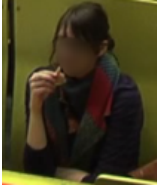
Image	Description
	The person looks to the front, sits with the head up in the basic position, with one/both hands under the desk, in a front facing posture, and without performing any body movements.
	The person (on the right) looks to the front, sits with the head in the basic position, in a front facing posture, and performs self-grooming movements.
	The person looks to the front, sits with the head up in the basic position, in a front facing posture, and is drinking or eating.
	The person looks to the front, sits with the head in the basic position, in a front facing posture, and is moving with low visibility under the desk.
	The person looks to the front, sits with the head in the basic position, with both hands on the desk, in a front facing posture, and without performing any body movements.

Table 5.1

Attentive nonverbal behavior combinations - first 10 in descending order of frequency - continued

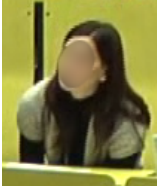| Image	Description
	The person looks to the front, sits with the head in the basic position, in a front facing posture, with an unclear hand posture, and without performing any body movements.
	The person looks to the front, sits with the head in the basic position, and is ordering objects on the desk.
	The person (on the right) looks to the front, sits with the head in the basic position, with one/both hands under the desk, in a front facing posture, and is having a conversation while looking to the front.
	The person looks to the front, sits with the head in the basic position, holding on an object, in a front facing posture, and without performing any body movements.
	The person looks to the front, sits with the head in the basic position, moving their hands, in a front facing posture, and is unpacking objects/packing up.

Inattentiveness (Table 5.2) was interpreted as a behavior where the coded person looks anywhere else but towards the speaker. In some cases, the coded person decides to do something else, like check their mobile phone, or unpack something from their bag. In others, someone else near the coded person does something that gets their attention. Actions that disturb a person and make them look away from the lecturer were called "action triggers".

Table 5.2

Inattentive nonverbal behavior combinations - first 10 in descending order of frequency

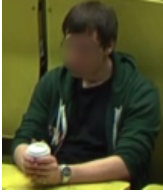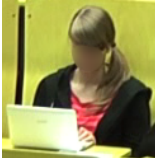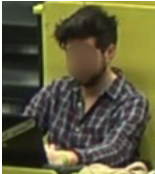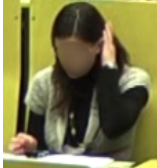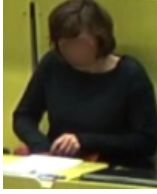| Image	Description
	The person looks down to the desk, sits with the head in the basic position, in a front facing posture, and is typing on a mobile device.
	The person looks down to the desk, sits with the head in the basic position, with one/both hands under the desk, in a front facing posture, and without performing any body movements.
	The person looks down to the desk, sits with the head in the basic position, in a front facing posture, without performing any body movements.
	The person looks left/right, sits with the head in the basic position, with one/both hands under the desk, in a front facing posture, and without performing any body movements.
	The person looks down to the desk, sits with the head in the basic position, in a front facing posture, and is moving with low visibility under the desk.

Table 5.2

Inattentive nonverbal behavior combinations - first 10 in descending order of frequency - continued

Image	Description
	The person (on the left) looks left/right, head in the basic position, one/both hands under the desk, in a front facing posture, having a conversation with a fellow student.
	The person looks down to the desk, sits with the head in the basic position, in a front facing posture, and is ordering objects on the desk.
	The person (on the left) looks left/right, sits with the head in the basic position, in a front facing posture, gesturing, speaking with a fellow student.
	The person looks down to the desk, sits with the head in the basic position, in a front facing posture, and is self grooming/scratching.
	The person looks down to the desk, sits with the head in the basic position, in a front facing posture, and is unpacking objects/packing up.

Below is a list of action triggers selected from the video recording:

- a person manipulating objects on the desk (minute 03:17, camera 2);

- a person interrupting the coded person to show her something on the phone (minute 02:10, camera 2);

- a person coming in late/leaving before class is over (minute 39:32, camera 2);

- a person watching something on their laptop (minute 19:06, camera 2).

The following list summarized virtual human features that can be further used to customize virtual characters and audiences:

1. Role. It can be used to create group dynamic behaviors through listener-listener interaction:

   (a) the action trigger for action initiator distracts attention of other virtual humans, creates rumor, restlessness, initiates conversations, etc.;

   (b) the follower for action recipient responds to the action trigger, becomes inattentive, restless, engages in conversations, etc.

2. Proportion. Technology experts suggested proportions to establish audience behaviors (e.g., the majority is attentive):

   (a) Proportions of behaviors across the whole audience: establish the proportion of various behaviors across the audience;

   (b) Proportions of people displaying the similar behavior across the audience: establish types of behaviors that the majority of people display at a given time).

3. Randomization before, during and after the speech. Technology experts suggested various ways to randomize behaviors:

   (a) by time of occurrence: use same behavior, chose different times of occurrence;

(b) by person: use same behavior, chose different person;

(c) by behavior: use same person, chose different behaviors;

(d) by intensity: use same behavior, chose different displayed intensity.

4. Frequency of change. According to the observation, the behavior that changed most frequently was gaze/head direction, followed by several complex body movements (e.g., actuating phones, unpacking objects) and hand and arm positions (between static and movement). Behaviors that changed least were body postures (between front and sideways) and nodding/head shaking.

   Taking this into account, a virtual human can be devised with a set of behaviors that can be varied at different intervals.

5. Behavior complexity. Attentiveness and inattentiveness displays can cover behaviors from static and simple to complex ones that involve many modalities with movements. In the observed sample, the majority of behavior combinations contained six behavioral categories that always included action trigger (speaker, self, or fellow student), eye gaze/head direction, head posture, and body posture plus any two others from the remaining categories (hand movements, agreement/disagreement, complex body movements, and conversations). Behaviors that are at the extremes, too simple or too complex were rare. Virtual humans could be animated with an average of six modalities that reflect the number observed with the real audience.

6. Activation. Attentiveness and inattentiveness were observed to have different activation levels. When people look to the front, they perform less hand and complex body movements, than when they look elsewhere. Therefore, virtual humans could be devised to be more active when they turn their gaze away from the speaker and perform less movements when they look towards the performer.

7. Duration. Behavior duration was calculated in the observation study and, on average, attentiveness lasted for six seconds and inattentiveness for eight seconds.

Some of the presented behaviors have been implemented in the CAVE-like environment at TU Ilmenau, within an early version of a virtual

audience. The application is part of a dissertation focused on the animation of virtual humans and on the creation of a graphic user interface. Figure 5.1 shows two examples of virtual humans within the application who embody attentiveness and inattentiveness.
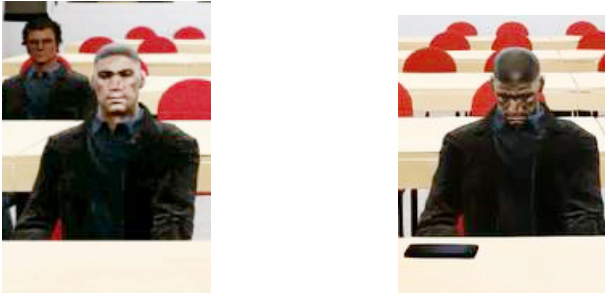
Figure 5.1. Virtual audience members developed at TU Ilmenau

5.4.1.3 Virtual room

The space occupied by a virtual audience is an indirect component that appeared in the interview analysis. The list of virtual rooms is build on the discussions with experts in the interviews and on literature on communication in the discipline approach. Locations should be adapted to the needs of the trainees and depending on the places where they are likely to deliver their presentations and speeches. The locations differ in terms of seating organization and audience size. Also, depending on the setting, the speaker can also stand at the same level with the audience (e.g., in a seminar room), on a higher position than the audience (e.g., in a conference hall), or on a lower position (e.g., in a lecture hall). Future studies could be conducted to see if and how this positioning impacts on performance. Three possible locations emerged as relevant:

1. A usual seminar room at the university, with up to 30 seats (as the rooms at Technische Universität Ilmenau). Tables are arranged on rows. The speaker stands in front of the room. This setting can be used for usual seminar talks or conference presentations. This configuration was used in the speech practice session observed in the first study and is presented in Figure 5.2.
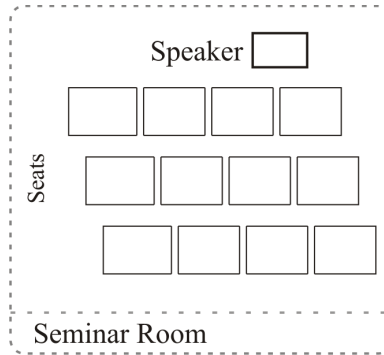
Figure 5.2. Proposed seminar room format

2. A conference room (Figure 5.3) with tables arranged in a circular form, with up to ten seats. People sit around the table. The speaker stands in front of the room. This setting can be used for job interviews, oral examinations, or business pitches.
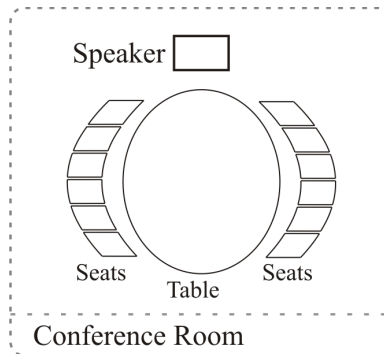
Figure 5.3. Proposed conference room format

3. A lecture hall at the university or a conference hall with more than 100 seats can be arranged in a semicircle, or in parallel rows having the speaker stand on a stage (Figure 5.4). This setting can be used for conference presentations and lectures presentations, and it was mentioned by one of the trainers in the setting of public

speaking competitions.
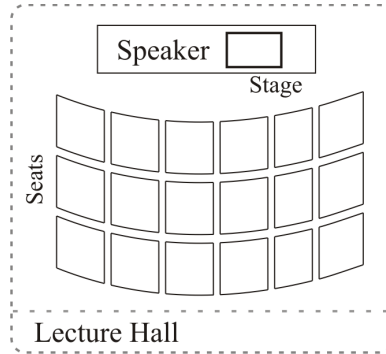
Figure 5.4. Proposed lecture hall format

Location features:

1. Place assignment: assign persons with certain characteristics and behaviors in key positions.

2. Seat occupancy: assign seat occupancy from 0-100% percent.

The next section will review the concepts for graphic user interface, or the options experts would like to have in order to customize the audiences and control them.

5.4.1.4 Suggestions for a graphic user interface

The last element reported here summarizes the expectations and suggestions of interviewees with regard to setting up and controlling a virtual audience. These options can be divided into three categories: before, during, and after the speech in front of the virtual audience. All three belong to the second level of the audience model because they imply a trainer's interaction with the virtual audience. At procedural level, the GUI contains the option to select features and behaviors as well as the option to randomize them across the virtual audience. Randomization was initially intended for the virtual humans behaviors, but it is sensible to use it as an option to vary virtual human's features

and feedback characteristics as well. The list here is a summary of their concrete suggestions:

Before the speech:

- set virtual humans' looks and behaviors before the speech (when selecting a speech scenario) with options for

  - physical characteristics;
  - outfit;
  - type of behavior;
  - time of occurrence;
  - order of appearance;
  - intensity of behaviors;
  - idiosyncratic display.

- set between several options for audience size, location size, and place assignment

- set duration of simulation, in accordance to the speech duration

- set randomization rules

  - first assign *each person* a characteristic and a behavior, than randomize all persons in space and time (bottom-up variety); *OR*
  - first assign characteristics and behaviors to the *whole group*, than change their proportions among virtual humans (top-down variety).

During the speech

- change virtual humans' behaviors during the speech (synchronous control of audience) with options for

  - attentiveness;
  - inattentiveness.

- choose feedback duration

- choose feedback type

    - related to presentation content;
    - related to presentation style.

- set randomization rules for who and when gives the feedback

After the speech

- choose feedback duration

- choose feedback type

    - related to presentation content;
    - related to presentation style.

- set randomization rules for who and when gives the feedback

- save user profile for further training sessions

In order to be able to use these options, experts wish to have a simple interface that can be maneuvered on the spot, especially during the speech practice. It should contain preferably drop-down lists and buttons, that allow an easy setup and audience control.

5.4.2 Audience scenario

Out of the 14 coded persons in the audience observation study five were randomly selected, and a scenario of five minutes is proposed. Each table contains the behaviors of one person. For this scenario, five minutes of coding were selected (the five minutes coded in the middle of the recorded lecture (minute 16.30-21.30) and the observed behaviors were transcribed in the given tables. Their order follows the exact order of the behaviors displayed by the listeners.

Each line represents a behavior combination at a certain moment in time. In the first left column of the table, there is the number of the coded person. The next column represents the behavior duration, given in seconds (s). In the "Gaze direction" column, "down" can take several values, such as "to desk/electronic devices/fellow student in the

front". The fifth column represents agreement or disagreement (A/D) behaviors, such as head nods or head shakes. The cells contain the textual translation of the behavior codes from the original code sheet.

This detailed approach has the advantage that one can see exactly which nonverbal modalities are active in each behavior combination. Second, one can see how long each behavior combination lasts. The fluidity of behaviors (which modalities shift faster than others) can also be observed here, as it was discussed in the results chapter of the audience observation study.

Table 5.3
Scenario proposal - Person 1

Duration (s)	Action trigger	Gaze direction	A/D	Head position	Hand and arm position	Body orientation	Body movements	Conversation
18	self	down	none	basic	moving their hands	front-facing	typing on mobile device/laptop	no
3	speaker	to the front	none	basic	holding an object	front-facing	self-grooming/scratching	no
46	self	down	none	basic	moving their hands	front-facing	typing on mobile device/laptop	no
4	speaker	to the front	none	basic	moving their hands	front-facing	self-grooming/scratching	no
3	self	down	none	basic	moving their hands	front-facing	typing on mobile device/laptop	no
1	speaker	to the front	none	basic	one/both hands under table	front-facing	no body movements	no
1	self	down	none	basic	one/both hands under the table	front-facing	no body movements	no
2	speaker	to the front	none	basic	one/both hands under the table	front-facing	no body movements	no

Person 1

Table 5.3
Scenario proposal - Person 1- continued

	Duration (s)	Action trigger	Gaze direction	A/D	Head position	Hand and arm position	Body orientation	Body movements	Conversation
Person 1	176	self	down	none	basic	moving their hands	front-facing	typing on mobile device/laptop	no
	2	speaker	to the front	none	basic	holding an object	front-facing	no body movements	no
	38	self	looking down	none	basic	moving their hands	front-facing	typing on mobile device/laptop	no

Table 5.4
Scenario proposal - Person 2

	Duration (s)	Action trigger	Gaze direction	A/D	Head position	Hand and arm position	Body orientation	Body movements	Conversation
Person 2	2	speaker	to the front	none	basic position	one/both hands under the table	front-facing	no body movements	no
	3	self	down	none	basic position	one/both hands under the table	front-facing	no body movements	no
	1	speaker	to the front	none	basic position	one/both hands under the table	front-facing	no body movements	no
	17	speaker	to the front	none	basic position	moving their hands	front-facing	drinking/eating	no
	2	self	down	none	basic position	one/both hands under the table	front-facing	no body movements	no
	1	speaker	to the front	none	basic position	moving their hands	front-facing	drinking/eating	no
	5	self	down	none	basic position	moving their hands	front-facing	drinking/eating	no
	4	speaker	to the front	none	basic position	moving their hands	front-facing	drinking/eating	no

Table 5.4
Scenario proposal - Person 2- continued

	Duration (s)	Action trigger	Gaze direction	A/D	Head position	Hand and arm position	Body orientation	Body movements	Conversation
Person 2	7	self	down	none	basic position	moving their hands	front-facing	ordering objects	no
	1	speaker	to the front	none	basic position	holding an object	front-facing	no body movements	no
	10	self	looking down	none	basic position	not identified	front-facing	drinking/eating	no
	12	speaker	to the front	none	basic position	moving their hands	front-facing	self-grooming/scratching	no
	7	self	down	none	basic position	moving their hands	front-facing	no body movements	no
	7	speaker	to the front	none	basic position	other	front-facing	no body movements	no
	35	self	down	none	basic position	one/both hands under the table	front-facing	taking notes	no
	2	other	left/right	none	basic position	one/both hands under the table	front-facing	drinking/eating	no

Table 5.4
Scenario proposal - Person 2- continued

Duration (s)	Action trigger	Gaze direction	A/D	Head position	Hand and arm position	Body orientation	Body movements	Conversation
6	self	down	none	basic position	moving their hands	front-facing	drinking/eating	no
1	other	left/right	none	basic position	one/both hands under the table	front-facing	drinking/eating	yes
8	self	down	none	basic position	one/both hands under the table	front-facing	drinking/eating	no
9	speaker	to the front	none	basic position	one/both hands under the table	front-facing	drinking/eating	no
10	self	down	none	basic position	moving their hands	front-facing	drinking/eating	no
22	speaker	to the front	none	basic position	one/both hands under the table	front-facing	drinking/eating	no
5	self	down	none	basic position	moving their hands	front-facing	drinking/eating	no

Person 2

Table 5.5
Scenario proposal - Person 3

Duration (s)	Action trigger	Gaze direction	A/D	Head position	Hand and arm position	Body orientation	Body movements	Conversation
47	self	down	none	basic position	one/both hands under the table	front-facing	no body movements	no
4	speaker	to the front	none	basic position	one/both hands under the table	front-facing	no body movements	no
67	self	down	none	basic position	one/both hands under the table	front-facing	no body movements	no
1	speaker	to the front	none	basic position	one/both hands under the table	front-facing	no body movements	no
12	self	down	none	basic position	one/both hands under the table	front-facing	no body movements	no
3	speaker	to the front	none	basic position	one/both hands under the table	front-facing	no body movements	no

Person 3

Table 5.5
Scenario proposal - Person 3- continued

	Duration (s)	Action trigger	Gaze direction	A/D	Head position	Hand and arm position	Body orienta-tion	Body movements	Conver-sation
Person 3	19	self	down	none	basic position	one/both hands under the table	front-facing	no body movements	no
	1	self	left/right	none	basic position	one/both hands under the table	front-facing	no body movements	no
	8	self	down	none	basic position	one/both hands under the table	front-facing	no body movements	no
	2	speaker	to the front	none	basic position	one/both hands under the table	front-facing	no body movements	no
	36	self	down	none	basic position	one/both hands under the table	front-facing	no body movements	no
	2	speaker	to the front	none	basic position	moving their hands	front-facing	self-grooming/ scratching	no

Table 5.5
Scenario proposal - Person 3- continued

	Duration (s)	Action trigger	Gaze direction	A/D	Head position	Hand and arm position	Body orientation	Body movements	Conversation
Person 3	3	self	down	none	basic position	one/both hands under the table	front-facing	moving with low visibility under the desk	no
	1	self	left/right	none	basic position	one/both hands under the table	front-facing	no body movements	no
	68	self	down	none	basic position	one/both hands under the table	front-facing	no body movements	no
	2	speaker	to the front	none	basic position	moving their hands	front-facing	self-grooming/scratching	no
	17	self	down	none	basic position	one/both hands under the table	front-facing	no body movements	no

Table 5.6
Scenario proposal - Person 4

Duration (s)	Action trigger	Gaze direction	A/D	Head position	Hand and arm position	Body orientation	Body movements	Conversation
110	speaker	to the front	none	head on palm(s)	other	front-facing	no body movements	no
1	self	down	none	basic position	one or both hands under the table	front-facing	no body movements	no
6	speaker	to the front	none	basic position	one or both hands under the table	front-facing	no body movements	no
8	self	down	none	basic position	one or both hands under the table	front-facing	taking notes	no
1	speaker	to the front	none	basic position	one or both hands under the table	front-facing	no body movements	no
43	speaker	to the front	none	basic position	both hands on table	front-facing	playing with objects/doodling	no
10	speaker	to the front	none	basic position	one/both hands under the table	front-facing	no body movements	no

Person 4

Table 5.6
Scenario proposal - Person 4- continued

Duration (s)	Action trigger	Gaze direction	A/D	Head position	Hand and arm position	Body orientation	Body movements	Conversation
11	self	left/right	none	basic position	one/both hands under the table	front-facing	no body movements	no
8	speaker	to the front	none	basic position	one/both hands under the table	front-facing	no body movements	no
3	other	left/right	none	basic position	one/both hands under the table	front-facing	no body movements	no
39	speaker	to the front	none	basic position	one/both hands under the table	front-facing	no body movements	no
4	self	left/right	none	basic position	one/both hands under the table	front-facing	no body movements	no
38	speaker	to the front	none	basic position	one/both hands under the table	front-facing	no body movements	no

Person 4

Table 5.7
Scenario proposal - Person 5

Duration (s)	Action trigger	Gaze direction	A/D	Head position	Hand and arm position	Body orientation	Body movements	Conversation
7	speaker	to the front	none	head on palm(s)	other	front-facing	taking notes	no
1	self	down	none	basic position	both hands on table	front-facing	taking notes	no
8	speaker	to the front	none	basic position	both hands on table	front-facing	taking notes	no
9	self	down	none	basic position	both hands on table	front-facing	taking notes	no
1	speaker	to the front	none	basic position	both hands on table	front-facing	taking notes	no
1	self	left/right	none	basic position	both hands on table	front-facing	taking notes	no
1	speaker	to the front	none	basic position	both hands on table	front-facing	taking notes	no
9	self	down	none	basic position	moving their hands	front-facing	self-grooming/scratching	no
60	speaker	to the front	none	head on palm(s)	other	front-facing	no body movements	no

Person 5

Table 5.7
Scenario proposal - Person 5- continued

Person 5

Duration (s)	Action trigger	Gaze direction	A/D	Head position	Hand and arm position	Body orientation	Body movements	Conversation
18	self	down	none	basic position	moving their hands	front-facing	drinking/eating	no
7	speaker	to the front	none	head on palm(s)	other	front-facing	no body movements	no
5	self	down	none	basic position	both hands on table	front-facing	no body movements	no
41	self	left/right	none	basic position	moving their hands	front-facing	moving with low visibility under the desk	no
5	speaker	to the front	none	basic position	one/both hands under the table	front-facing	moving with low visibility under the desk	no
4	self	looking down	none	basic position	one/both hands under the table	front-facing	self-grooming/scratching	no
110	speaker	to the front	none	basic position	one/both hands under the table	front-facing	no body movements	no

Table 5.8 illustrates an exemplification of audience characteristics for
each of the persons chosen in the scenario above. The column titles
represent the categories provided in the audience concept lists, except
for status and occupation that were merged here in one single column.
All values were chosen randomly. Features such as place of origin,
religious affiliation, political views, and hobbies would have to be
discussed first with the speaker, before establishing the similarities
and differences to their characteristics, and before devising virtual
characters that act upon them.

Table 5.8
Character description

Audience member	Known to speaker	Age	Sex	Status and occupation	Place of origin (country)	Outfit	Religious affiliation	Political views	Hobbies
Person 1	yes	22-34	F	same (student)	same	casual	similar	different	different
Person 2	yes	45-54	M	different (managerial position)	same	casual	different	different	similar
Person 3	no	22-34	M	different (non-manage-rial position)	different	event-based	different	similar	similar
Person 4	no	22-34	M	same (student)	same	event-based	similar	similar	different
Person 5	no	55-64	F	different (managerial position)	different	casual	different	different	different

5.4.3   Summary

The audience concept and scenario summarize all findings of the empirical studies. These were reformulated from the expert interviews, speech practice observation sheets, and student lecture observation lists and rearranged into a simplified version of a game design document and scenario tables. The audience characteristics given here represent the perspective on a real audience member (as recalled by experts in their work experience or observed in the student lecture). The ideal technical attributes described by (Kasap & Magnenat-Thalmann, 2007) were not addressed separately in the audience concept, but one can deduce the level of artificial intelligence needed to achieve virtual audiences that embody the wide variety of human-like attributes contained in the concept.

The behaviors presented in the scenario are only a snapshot of those identified in the student audience observational study. Starting from this scenario, one can create an audience prototype by endowing each of the five audience members with the characteristics in the audience concept. Table 5.8 is an example of characteristics embodied by the virtual humans. Furthermore, other scenarios can as well be created from the one proposed here, by applying randomization rules on behavior durations (column 2 in the scenario tables) in order to vary manifested attentiveness and inattentiveness towards a speaker.

CHAPTER 5.5

Conclusion

The design chapter presented the audience concept and a scenario proposal based on the previous empirical studies. Although the data collected for the design concept is grounded on theories and methodologies applied in social sciences, it had to be reorganized in line with the virtual environments design literature. Following the literature review, a general conclusion emerged and it regarded the little consensus on the fittest theories and methods for virtual audience design. This aspect became even more visible for virtual environments planned for educational purposes in a domain so complex like human communication. A few theoretical studies in the educational field (Delia, 2003; F. Mantovani & Castelnuovo, 2003; Sánchez et al., 2000) though provided enough support for the creation of the audience concept and scenario. However, the suggestions given here with regard to the presentation of the audience concept in the form of lists are not normative. It is believed that once the audience prototype is created, future studies would help improve not only the animated content, but also the design process and offer clearer paths to follow.

The lists in the audience concept section summarized the findings of both studies and presented several features that an audience concept should address in the design process. Most notable findings included audience demographic characteristics which should allow using a large pool of potential virtual audience members of different ages, looks, race, religions, and occupations. With regard to their behaviors, they could embody any of the 225 different nonverbal displays of attentiveness and inattentiveness identified in the observation study. These features can be further tested through usability studies.

The scenario section represents an instant of the observed student audience, which could serve as model for the virtual one. Although the scenario contains only descriptions of behaviors, the characters could embody any of the physical characteristics presented in the audience

concept and could be subject to any setup and control rules given by trainers.

A through discussion of the audience concept and scenario is given in the general discussion chapter of the dissertation, where all the findings of the empirical studied are put into perspective through the research model suggested at the beginning of the dissertation.

6

General discussion

> *If you're not comfortable with public speaking - and nobody*
> *starts out comfortable; you have to learn how to be comfortable*
> *- practice. I cannot overstate the importance of practicing. Get*
> *some friends or family members to help evaluate you, or*
> *somebody at work that you trust.*

<div align="right">Hillary Clinton</div>

Public speaking is an activity that everyone might have to perform at one point or another in their life. However not everybody feels comfortable speaking to a crowd and for many people this experience can be so unpleasant that they avoid it altogether. To address this problem, speakers can undergo treatment to learn how to manage their anxiety levels, or they can join communication skills courses that teach them how to improve their speech performance. As seen in in the observation study of a speech practice session, during public speaking seminars, participants are introduced into the theoretical background in communication and have the opportunity to practice speeches in front of their fellow seminar participants. What if though, apart from these usual listeners one would have access to different audiences, that look and behave diversely, are not judgmental, unless set so, and are available for training at any possible time?

The present dissertation proposes a training tool in the shape of a virtual audience that can be used to specifically address speech skills improvement. The literature review on existing virtual audiences revealed that they have been constructed to test whether they could work as treatment tool (North et al., 1998; Wallach et al., 2009) and actually help people overcome speech anxiety in clinical studies (H. V. Anderson, 2003; Wallach et al., 2011). Both endeavors proved successful. However, groups of virtual characters haven't been used so far outside the laboratory, to train someone master proper public speaking. In the communication training field, the tool of choice has been so far a singular virtual human who simulates virtual patients and trains physicians and therapists how to interview and diagnose patients (Cook & Triola, 2009; P. Kenny et al., 2007; Pataki et al., 2012; Saleh, 2010). With such sustained attention towards virtual characters in education, a virtual audience for public speaking and presentation skills training appeared as a legitimate next step of research. Not only could

this address the interests of various market segments, but would also push virtual reality solutions out of the laboratory and into the training curricula of education institutions. This view is further supported by the role public speaking and presentation skills courses occupy in higher education curricula, where their main role is to prepare students to communicate properly in school and at their jobs (Dunbar et al., 2006; Plutsky, 1996). The aspect is particularly important when employees have to present in front of clients and of people from other study or work backgrounds. Dannels (2001) and Dannels and Housley Gaffney (2009) underline this when they call for a cross-disciplinary approach to communication skills to prepare people for diverse audiences and business environments.

Immersive virtual reality technologies have been undergoing constant development since Ivan Sutherland's head-mounted three dimensional display from 1968 (Sutherland, 1968) and are preparing to hit the market with the much awaited entertainment HMDs from Oculus Rift, MergeVR, or HTC Vive (Gaudiosi, 2015). Apart from reaching new stakeholder groups in the entertainment industry, virtual reality can be further adapted to expand beyond natural science and mathematics teaching (Mikropoulos & Natsis, 2011) towards communication training education.

In order to populate an immersive virtual environment with believable virtual audiences and characters at 1:1 scale that behave realistically, it's important first to understand the needs of those who would work with them there and to know what virtual humans ought to do. The present dissertation was guided by one main question: What features describe an audience and what do people do when they sit and listen? Expert interviews and two observational studies were conducted to answer this question, and their findings were used to construct a virtual audience concept proposal and an actual scenario with five audience members that can be animated in virtual reality.

CHAPTER 6.1

The research model for a virtual audience

Based on this state of research and theoretical background, a research model for a virtual audience in communication training was proposed at the beginning of the present research. Figure 6.1 contains a diagram with revisions added to the core elements–audience, interactivity with the speaker, and training context as a situated usage of the virtual audience within the training program. Next to these initial elements are those that emerged from the empirical studies.
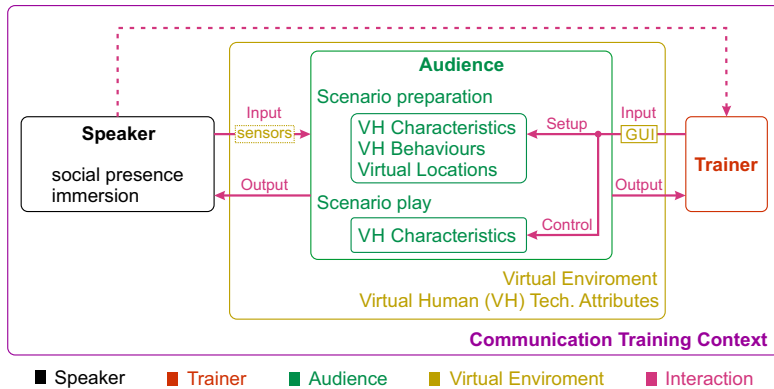

Figure 6.1. Research model for a virtual audience

Due to the complex findings, the structure of the model was revised and the next section will detail each implemented change and further explain the model.

6.1.1 Audience

The audience level (marked with "Audience" in green) kept its initial centrality in the research. In the graph it can be observed though, that the virtual audience used for training is represented as a comprehensive scenario containing not only its characteristics, but also behaviors, and the locations it can inhabit.

Empirical findings showed that trainers wish to customize their audiences with various characteristics and behaviors that are relevant for the training purpose of each speech seminar participant. They also have an active role in guiding the feedback session of the audience at the end of each speech practice (public speaking seminar observation) and expect ample feedback on the speech from the virtual audience as well.

There is also great potential for audience variety based on age, outfit, occupation, status, and religious background. Moreover, virtual humans could take up more complex features such as looking familiar and unfamiliar to speakers. Findings show that sometimes speakers benefit more if they speak to audiences they know, because they are supportive and people enjoy a friendly atmosphere during training. However, in order to improve skills and performance, speakers have to be exposed to new faces as well, and a virtual audience that looks totally unfamiliar could fulfill this role. As the seminar progresses, unfamiliar faces may be more helpful than familiar ones.

These options for virtual humans go way beyond the usual features used so far in fear of public speaking training. Even if audiences were made to look differently (have different clothing and physical features, as in Grillon et al. (2006)), these features weren't used to train people to adapt their speech content and delivery style to the virtual characters according to their displayed age, social status, or the outfit they wore.

Apart from physical characteristics, audiences embody various behaviors that include not only attentive and inattentive ones identified in the observed group of students in relation to a speaker, but also group dynamic behaviors, in relation to other members of the audience. Following the literature on nonverbal communication (Hale & Burgoon, 1984; Miles & Johnston, 2007; Norton & Pettegrew, 1979), attentive and inattentive behaviors were differentiated based on gaze direction: when a person looks towards the speaker, it is coded as attentive; when

they look elsewhere, be it towards the desk or fellow student, they are coded as inattentive. Findings showed that people can display attentiveness and inattentiveness in extremely different manners. Even more, contrary to common sense expectations, they rarely sit still, even if they pay attention to the speaker. They play with pens or with mobile devices, pass on objects, eat, talk, unpack objects, and, in extreme cases, can perform all these behaviors simultaneously. Furthermore, other detailed findings such as behavior duration and the rate at which nonverbal modalities change (gaze changes most often, whereas position of head rarely varies) help establish animation priorities: which behaviors have to be animated more often than others.

Group dynamic analysis both in the first and in the second study revealed that there are people whose behaviors trigger reactions in others. For instance, as people would walk up and down the stairs to enter or leave the lecture hall (trigger), fellow students would gradually get distracted (follower) as the person would pass by them. Virtual humans can take up the role of trigger and follower as well, to simulate interaction among virtual groups. These findings are compatible with the spatial model of interaction in VR proposed by Benford and Fahlén (1993) and which describes principles for perception and interaction within groups of virtual characters. Dynamic audiences where listeners interact with each other are also very important from the training perspective, because, according to communication experts, speakers have to become habituated to restless audiences and learn management strategies to get their attention back. This marks an important difference between phobics and trainees: whereas anxious speakers are taught how to lower their anxiety levels when talking to an inattentive virtual audience (Wallach et al., 2009), trainees are expected to learn how to grab the attention of listeners who lost their interest.

Throughout the interview with training experts, audience location also came up. Speech coaches recalled several training situations, and a few potential training locations could be inferred. The animation of virtual humans is closely related to the virtual spaces they inhabit and 3D animation can be planned based on obstacles (e.g., furniture) and distance from the speaker (e.g., seat allocation relative to the speaker's position). Stuevel et al. (2014) suggests using 2D, simplified impostors for characters that are far away from the observer and which don't need too much visual detail. Simplified animations can as well be implemented for virtual humans that are seated and who

are only visible from torso up.   Due to this close relation between space allocation and visible VH features and animations, the virtual room suggestions ("Virtual Locations", marked in green in the research model) are presented also as part of audience features.

Audience characteristics, behaviors, and occupied locations can form concrete training scenarios (marked in the diagram with "Scenario"). Although the scenario presented in the previous chapter contains only the behaviors of five audience members, any of those can take up any physical or cultural characteristics proposed in the design concept and occupy any desirable virtual room.   Audience characteristics and behaviors customizations as well as realistic representations through scenarios bear also theoretical relevance, because these elements support the constructivist view on a learner's diverse needs.

6.1.2   Virtual environments

This is a new level (marked with "Virtual Environment" in yellow in Figure 6.1) that was added to the initial research model, and it contains elements related to AI ("VH Technical Attributes" marked in yellow) and to interaction technologies ("sensors" and "GUI" marked in yellow).

At first, the research model integrated technical attributes of autonomous intelligent virtual humans, such as autonomy, interactivity, and personification in the Audience level.   These were presented as desirable features for virtual humans who perceive their environments and the user's input, take decision based on that, and act on emotions and personality.   The empirical research helped clarify some of these features and how they can be met.   Observed group dynamic exemplified how people perceive their peers and how they react (e.g., through involuntary attention) based on what others do around them, to set examples for VH autonomous behaviors.   Attentive and inattentive behaviors exemplified how people act on their decision (to pay attention to the speaker or not), to set examples for VH interactivity.   Emotions and personality were not explicitly studied, but there are already many emotion models that can be used to add affective layers to virtual humans (e.g., the OCEAN and ALMA models (Kasap & Magnenat-Thalmann, 2007)).

AI attributes and interaction technologies are presented here in the same level together, because the sensors the system affords as well as its

control by the users are developed and implemented in close connection to the supported artificial intelligence of the system (desirable audience technical attributes).

6.1.3 Interaction

Interaction between speaker and audience was allocated in the initial research model to the second level. The revised model dropped the level denomination and shows interaction in purple and in the form of Input and Output on the side of the speaker (in black) and of the trainers (in red), since both of them interact with the virtual audience. The revised research model acknowledged the important role played by the trainer in the speech practice, therefore both trainer and trainee become equal partners on each side of the virtual audience in the interaction process. The dotted line from speaker to trainer marks also the potential interaction between these actors themselves during the speech practice in front of the virtual audience: the trainer observes how the speaker is doing and adjusts the virtual audience accordingly.

The first study dealt with interaction between speaker and listeners in real life settings and also with aspects of human-computer interaction. The most relevant finding regarding speaker-listener interaction is that real training situations allow audiences to participate with comments and questions on speech content and presentation in the feedback session at the end of each speech practice. According to the training experts, such complex feedback is desirable also from a virtual audience. Since audiences respond to both content and delivery style, virtual audiences are expected to have high artificial intelligence attributes (see previously explained audience technical characteristics).

From a technical perspective, interaction can be achieved through various sensors, and VR experts opted for less intrusive ones, like movement detectors (e.g., Kinect) and microphones (marked with "sensors" on the left side, in yellow). These sensors would pick up a speaker's verbal and nonverbal input and feed it into the system where virtual humans would ideally respond to it verbally and nonverbally. Sensors may not be obligatory (marked with an interrupted contour). However, in their absence the trainers would have to control virtual audience's responses.

In the diagram, both speakers and trainers act on the virtual audience through role-related inputs and outputs. The speaker has a verbal and nonverbal input during the speech (marked with "Input" on the left side), and becomes aware of the virtual audience's characteristics and behaviors through visual and aural cues, as the virtual audience is displayed in the immersive virtual environment (marked with "Output" on the left side).

The trainers act on the virtual audience in two different manners (marked with "Input" on the right side). They can set up the virtual audience ("Setup", in purple) before the speech ("Scenario preparation" marked in green in the audience level). Once the speech begins, trainers could control ("Control", in purple) the audience's reactions ("Scenario play", in green) such as attentiveness, inattentiveness, and group dynamic. The setup phase determines which values the virtual humans' features will take, what behaviors they can embody, and which locations they can populate. When the speech begins, the control phase can also start and this is where virtual audiences come to life under the supervision of the trainer. Speech begin does not automatically call for audience control. On the one hand, control depends on the level of artificial intelligence. If the audience is responsive to the speaker, it might need less or no control at all. If the virtual audience makes no use of sensors to interpret a speaker's verbal and nonverbal input, then trainers would have to steer its responses. On the other hand, control depends on the complexity of audience responses. Speech can be hard to simulate with AI, therefore, trainers could activate speech responses of prerecorded utterances. It is hypothesized that this interaction supports public speaking skill acquisition. The model suggests that also the trainer becomes aware of the virtual audience and its behaviors, by assisting, together with the speaker to the unfolding of the scenario (marked with "Output" on the right side, in purple).

Trainers suggested setting-up audiences and controlling them via interfaces with buttons, drop-down lists, accessible data bases of behaviors, as well as repositories with the progress history of each trainee. Technology experts stated as well that audiences should be controlled, instead of having autonomous ones, which are in their opinion hard to realize with the existing technologies. The interface ("GUI") becomes therefore an important tool in the proper functioning of a virtual audience, if not during the speech, at least in the scenario setup phases.

The highlight of the revised research model with regard to interaction is the fact that trainers becomes central actors in the virtual reality environment. In the end, following the CID paradigm, they become the customization experts for the needs of public speaking trainees who undergo virtual training.

The interactive capabilities presented here resonate with the model of presence in training environments proposed by F. Mantovani and Castelnuovo (2003). The researchers call for simulations with high graphic fidelity and the CAVE-like immersive environment suggested for the audience prototype fulfills these technical prerequisites. Control of the simulation is another topic that would influence experienced presence in learning environments and the GUI proposed here can serve this goal. Furthermore, the concept suggests ways to support cognitive-behavioral interaction between users and virtual characters, as well as social interaction. Through perception and interaction with virtual humans, users may experience social presence (G. Mantovani & Riva, 1999) in the company of virtual audience ("Social presence", in black). Last but not least, the fact that the virtual audience concept is based on real audience characteristics and behaviors demonstrates its commitment to the general ecological validity of the virtual content. Apart from the theoretical value of interactive virtual audiences, their feedback capabilities could help people improve questions and answers skills, which are so important in work settings (Johnson & Szczupakiewicz, 1987).

At this point the model raises new questions that can be discussed in future studies, such as the impact of audience setup and control, as well as trainer's presence during the simulation on experienced presence and their impact in speech improvement and effectiveness.

6.1.4 Communication training context

The third level of the audience model represents the larger communication training context where virtual audiences would operate (marked with violet). This level remained unchanged from the initial research model. The observation of a speech training session helped localize the role of speech practice within the economy of the training seminar. Speech practice is rather short, between 1-8 minutes, and takes place after trainers cover a theoretical part on communication. This is

in line with the practices described elsewhere (Beebe, 2007). For the particular observed speech training session, the speech practice session took place on the third (and last) day and was the last element of the seminar. Nevertheless, the feedback session for each participant who gave a presentation was extensive, even up to one hour per person.

Understanding how speech training seminars function could help devise a virtual audience that fits within the program duration and within the time slots allocated for practice. Additionally, the third level opens the discussion on the total time and effort trainers are expected to invest in the training, specifically for the customization and control of the virtual audience. This is particularly relevant, because, as people advance with their seminars and acquire new skills, complex customizations (e.g., with familiar or unfamiliar faces) might not be necessary anymore. Future studies could investigate how these activities impact seminar planning and how they interact with other training parts, such as theory learning or speech material preparation.

6.1.5   Audience design concept and scenario

The outcomes of the studies conducted here were synthesized into the virtual audience design concept and a suggested scenario. The design concept is based on a combination of design and educational theories that highlight the necessary characteristics of virtual learning environments. With the help of the game design document methods, relevant audience features identified in the empirical studies were brought together and described thoroughly.

The proposed scenario contains a snapshot of the observed student audience. Five coded persons were randomly selected and their behaviors were enlisted in the time order of their occurrence. The list contains the duration of each behavior combination and also the description of what movements the person was performing. The scenario can be implemented for the VA prototype at TU Ilmenau.

One could synthesize the findings of the studies and the audience concept into a few major ideas. First, training experts are enthusiastic about having a virtual audience as tool to prepare their pupils. When faced with questions about how to customize and control such an audience, one idea crystallized: simplicity. Both VR experts and trainers wish to have simple buttons and lists of behaviors that can be

easily exchanged when needed. However, none of the training experts had any previous experience with immersive virtual environments, which could influence their expectations regarding technology and its capabilities.

Another important aspect is focus on speech content. Public speaking training is heavily focused on feedback and this means having audiences that can ask questions related to the presentation and also give feedback on speech style and content. This raises the bar with regard to virtual audience capabilities, which should have enhanced artificial intelligence capabilities compared to the ones used in phobia treatment.

The third element is the level of customization. While in fear of public speaking the way audiences look, what they wear, their age, and status are elements of variety that make the audience realistic and close to a humane one, in public speaking they are tightly connected to how audience are expected to react to speech content and style. Therefore, in order to achieve a realistic training situation with optimum skill transfer, virtual audience characteristics should be relevant for the purpose of the speech content and speech situation.

CHAPTER 6.2

Limitations

One general limitation of research conducted for the audience concept is that it didn't include the opinions of trainees. In order to propose a final version of the audience, a complete stakeholder map should be developed and all potential users should be considered. Moreover, most of technology experts shared their experience or knowledge on phobia treatment projects that used virtual audiences. This proved disadvantageous, because it made it difficult to discuss with them what needs to be improved in the future. The examples they gave were less of imaginary nature (what would they wish to improve), but mirrored more the applications that are already on the market or that have been already extensively used in phobia treatment.

Another limitation addresses the proposed scenario. Although detailed with regard to behaviors, the scenario doesn't provide lists of objects used by listeners (e.g., pens, cups, bags, etc.) and these have to be designed separately and adapted to the described behaviors. The video recordings for the scenario can be used to extract some of the usual objects used by students in class. Moreover, other basic animations, such as breathing and blinking were also omitted, as well as any precise reference to events taking place outside the virtual room, which can be see through the window or overheard. The expert interviews gave some examples for the latter, and these can be implemented in the scenario at a later stage.

Regarding the virtual audience concept, only visual data was proposed for audience behaviors. Sounds were recorded in the analyzed video, but they were not coded. This issue can be overcome by sampling sounds separately and implementing them directly in virtual reality. Furthermore, the scenario contains behaviors displayed by student audiences. This could represent a bias towards a specific type of audience. The categories in the used codebook were taken from literature on student behaviors (e.g., Mann and Robinson (2009)) and

were pretested with a student audience. A more mature audience seated in a different setting than a lecture hall could behave differently and could require different categories in the codebook.

A limitation of the empirical studies lies in the small samples. For the expert interviews finding available experts who would have experience simultaneously with immersive virtual environments and virtual audiences proved rather difficult. For the audience observation, a bigger sample or a more compact group of students could have yielded a better understanding of group dynamic behaviors.

CHAPTER 6.3

Outlook

The first steps in creating a virtual audience have already been made at Technische Universität Ilmenau, where a virtual audience is currently being developed to serve as communication training tool in a CAVE-like virtual environment. Some of the behaviors identified in the audience observation study have been implemented as well. Among the next steps of research are the improvement of the user interface and sound implementation.

Other opportunities for future studies regards audiences in both training and treatment fields. The audience proposed here is intended to accommodate numerous detailed behaviors and be easily customizable to match training purposes. It would be interesting to know though if audiences used for phobia treatment that already exist could work for training purposes and whether simple exposure to an audience would determine performance improvement. A comparison of such audiences over different target groups would help define even better the particularities of a training audience and what characteristics can make it successful in training. One hypotheses is that, at basic level training, both audiences would prove equally successful. The more advanced the public speaking and presentation training becomes, the more specialized the audience and the higher the expectations on its capabilities could be. Therefore, the next hypothesis is that for advanced presentation and public speaking skills trainings, the virtual audiences should be highly interactive and be able to process both speech content and style.

Another future study could focus on the evaluation of the given scenario: does it seem plausible? Is it realistic? Do speakers feel present and what is the role of presence in public speaking training? The same could apply for the scenarios that combine attentive and inattentive behaviors: at what point is an attentive audience perceived as less attentive, and is there a threshold behind which people start

perceiving this change? For such a study, a measurement instrument has to be developed to enable people to evaluate attentiveness and inattentiveness within groups of virtual humans. The hypothesis is that perception changes when people display behaviors with higher intensity. The more fidgety people are, the more inattentive they can be perceived. The following hypothesis is that perception depends on numbers of affected people. A single inattentive person would not make a whole audience be regarded as inattentive, just as a single attentive one would not make a whole audience be seen as attentive. Therefore, a potential research question would be: How many people would it take to make an audience be perceived one way or another?

The interviews with technology experts opened important discussions about the complexity of a system that harbors such a virtual audience and its scenarios. At a first look it appears that the system should be highly complex to accommodate the wishes of trainers and the options for audience customization and control. How a user interface could look like and what is the computing power behind such versatile audiences remains to be researched in the future. Moreover, in order to create a viable virtual reality solution, experts from relevant fields should be able to participate together in the design process. Therefore, focus groups and task analyses could offer a deeper insight into what's relevant and doable both from a financial and a design perspective.

Another point that could be clarified in the future regards the control options of audiences. Trainers wish to have virtual audiences that can act on their own but that can also be controlled when needed, technology experts were rather skeptical with regard to autonomy. A few questions arise: What is the optimum control level for a virtual audience? How much control is needed in order to keep the audience realistic enough? How can trainers control an audience in real time by pressing buttons and remain focused on the speakers as well?

A further aspect that can be discussed with trainers in the future is to see in which way they would fit the virtual audience and the technology behind it into their practice. The observation study of a speech practice session identified the rather limited time assigned to speech practice and a virtual audience might be able to compensate that, by letting trainees practice on their own or outside the official training seminar hours. A research on how coaches would implement this technology into their schedule hints towards technology adoption and domestication theories

(E. Rogers, 2003). Results of such studies could further help simplify virtual audiences, scenarios, and their supporting technologies to fit the technical skills and adaptability of end users. If such training options prove successful, it would be interesting to find out if they can be used to evaluate performance next to the standardize methods, such as the NCA questionnaires. Furthermore, as public speaking training becomes a pervasive activity in many universities, the application that displays the virtual audience should also become easy to maneuver and to use, outside specialized laboratories and on more accessible and commercially available platforms, such as the head-mounted displays preparing to enter the market.

References

Allen, M., Hunter, J. E., & Donohue, W. (1989). Meta-analysis of self-report data on the effectiveness of public speaking anxiety treatment techniques. *Communication Education*. doi: 10.1080/03634528909378740

Alverson, D. C., Saiki, Jr., S. M., Caudell, T. P., Goldsmith, T., & Stevens, S. (2006). Reification of abstract concepts to improve comprehension using interactive virtual environments and a knowledge-based design: A Renal physiology model. In J. Westwood et al. (Eds.), *Medicine meets virtual reality 14. accelerating change in healthcare: Next medical toolkit.* IOS Press.

Andersen, J. F., Andersen, P. A., & Jensen, A. D. (1979). The measurement of nonverbal immediacy. *Journal of Applied Communication Research*. doi: 10.1080/00909887909365204

Anderson, H. V. (2003). A Virtual end to stage fright: Sophisticated new technology spells the end of podium panic. *Harvard Business Publishing Newsletters*, 3–5.

Anderson, P. L., Price, M., Edwards, S. M., Obasaju, M. a., Schmertz, S. K., Zimand, E., & Calamaras, M. R. (2013, October). Virtual reality exposure therapy for social anxiety disorder: A Randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *81*(5), 751–60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23796315 doi: 10.1037/a0033559

Anderson, P. L., Rothbaum, B. O., & Hodges, L. F. (2003). Virtual reality exposure in the treatment of social anxiety. *Cognitive and Behavioral Practice*, *10*(3), 240–247. Retrieved from http://dx.doi.org/10.1037/a0033559

Anderson, P. L., Zimand, E., Hodges, L. F., & Rothbaum, B. O. (2005, January). Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and Anxiety*, *22*(3), 156–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16231290 doi: 10.1002/da.20090

Ayres, J. (1996). Speech preparation processes and speech apprehension. *Communication Education*, *45*(3), 228–235. Retrieved

265

from `http://ezproxy.uakron.edu:2048/login?url=http://`
`search.ebscohost.com/login.aspx?direct=true{&}db=`
`eric{&}AN=EJ532171{&}site=ehost-live`

Bakthin, M. M. (1986). *Speech genres and other late essays*. University of Texas Press.

Baños, R. M., Botella, C., Rubió, I., Quero, S., García-Palacios, A., & Alcañiz, M. (2008). Presence and emotions in virtual environments: the influence of stereoscopy. *Cyberpsychology & Behavior*, *11*(1), 1–8. Retrieved from `http://www.liebertonline.com/doi/abs/` `10.1089/cpb.2007.9936$\delimiter"026E30F$nhttp://www` `.tandfonline.com/doi/abs/10.1080/07481180802231319$\` `delimiter"026E30F$nhttp://online.liebertpub` `.com/doi/abs/10.1089/cyber.2012.1572$\` `delimiter"026E30F$nhttp://online.liebertpub.com/doi/` `abs/10.1089/cpb.2007.9936` doi: 10.1089/cpb.2007.9936

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

Barkhuus, L., & Jørgensen, T. (2008). Engaging the crowd: Studies of audience-performer interaction. In *CHI 2008 Proceedings* (pp. 2925–2930). doi: 10.1145/1358628.1358785

Barsade, S. G. (2002, December). The Ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly*, *47*(4), 644. doi: 10.2307/3094912

Bartlett, E. E., Grayson, M., Barker, R., Levine, D. M., Goldern, A., & Libber, S. (1984). The Effects of physical communications skills on patient saatisfacetion: Recall and adherence. *Journal of Chronical Dissease*, *37*(9), 755–764.

Bassett, J., Marshall, P., & Spillane, R. (1987, December). The Physiological measurement of acute stress (public speaking) in bank employees. *International Journal of Psychophysiology*, *5*(4), 265–273. Retrieved from `http://linkinghub.elsevier.com/` `retrieve/pii/0167876087900584` doi: 10.1016/0167-8760(87) 90058-4

Beattie, G., & Shovelton, H. (1999, December). Mapping the range of information contained in the cconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology*, *18*(4), 438–462. Retrieved from `http://jls` `.sagepub.com/cgi/doi/10.1177/0261927X99018004005` doi:

10.1177/0261927X99018004005

Beatty, M. J. (1988). Situational and predispositional correlates of public speaking anxiety. *Communication Education*, *37*(1), 28–39. doi: 10.1080/03634528809378701

Beatty, M. J., Balfantz, G. L., & Kuwabara, A. Y. (1989). Trait-like qualities of selected variables assumed to be transient causes of performance state anxiety. *Communication Education*, *38*(3), 277–289. doi: 10.1080/03634528909378763

Beebe, S. A. (2007, apr). Raising the question #6: What do communication trainers do? *Communication Education*, *56*(2), 249–254. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/03634520601145266 doi: 10.1080/03634520601145266

Beebe, S. A., & Beebe, S. J. (2010). *A Concise public speaking handbook* (3rd ed.). Pearson.

Beebe, S. A., & Beebe, S. J. (2012). *Public speaking handbook* (4th ed.). New Jersey: Pearson Education.

Beebe, S. A., Mottet, T., & Roach, K. D. (2012). *Training and development: Communicating for success* (2nd ed.). Pearson.

Beier, K.-P. (2008). *Virtual reality: A Short introduction.* Retrieved from http://www-vrl.umich.edu/intro/

Beiseler, M. (2004). *Akademisches Klopfen.* Retrieved from http://dw.com/p/4zc0

Benford, S., & Fahlén, L. (1993). A Spatial model of interaction in large virtual environments. In G. De Michelis, C. Simone, & K. Schmidt (Eds.), *Proceedings of the Third European Conference on Computer-Supported Cooperative Work* (pp. 109–124). Milan.

Bickmore, T., & Giorgino, T. (2006). Health dialog systems for patients and consumers. *Journal of Biomedical Informatics*, *39*(5), 556–571. doi: 10.1016/j.jbi.2005.12.004

Biocca, F. (1997, jun). The Cyborgs dilemma: Progressive embodiment in virtual environments. *Journal of Computer-Mediated Communication*, *3*(2), 1–17. Retrieved from http://doi.wiley.com/10.1111/j.1083-6101.1997.tb00070.x doi: 10.1111/j.1083-6101.1997.tb00070.x

Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators and Virtual Environments*, *12*(5), 456–480. doi: 10.1162/105474603322761270

Blanchard, A., & Markus, M. (2002). Sense of virtual community: Maintaining the experience of belonging. In *Proceedings of the*

35th Hawaii International Conference on System Sciences (pp. 3566–3575). IEEE Comput. Soc. doi: 10.1109/HICSS.2002 .994449

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*(2), 103–124. Retrieved from `http://vhil.stanford.edu/pubs/2002/blascovich-IVET.pdf`

Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4th ed.). Heidelberg: Springer.

Botella, C., Gallego, M., Garcia-Palacios, A., Guillen, V., Baños, R., Quero, S., & Alcañiz, M. (2010). An Internet-based self-help treatment for fear of public speaking: A Controlled trial. *Cyberpsychology, Behavior, and Social Networking*, *13*(4), 407–421. Retrieved from `http://www.liebertonline.com/doi/abs/10.1089/cyber.2009.0224` doi: 10.1089/cyber.2009.0224

Botella, C., Gracía-Palacio, A., Villa, H., Baños, R., Quero, S., Alcañiz, M., & Riva, G. (2007). Virtual reality exposure in the treatment of panic disorder and agoraphobia: A Controlled study. *Clinical Psychology and Psychotherapy*, *14*(3), 164–175. Retrieved from `http://www.embase.com/search/results?subaction=viewrecord{&}from=export{&}id=L46951842$\delimiter"026E30F$nhttp://dx.doi.org/10.1002/cpp.524`

Botella, C., Villa, H., García-Palacios, A., Baños, R. M., & Perpiñá, C. (2004). Clinically significant virtual environments for the treatment of panic disorder and agoraphobia. *Cyberpsychology & Behavior*, *7*(5). doi: 10.1089/cpb.2004.7.527

Bouchard, S., St-Jacques, J., Robillard, G., & Renaud, P. (2008). Anxiety increases the feeling of presence in virtual reality. *Presence*, *17*(4), 376–391.

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., Gabbard, J. L., & Hix, D. (2002). A Survey of usability evaluation in virtual environments: Classification and comparison of methods. *Teleoperators and Virtual Environments*, *11*(4), 404–424.

Bowman, D. A., & McMahan, R. P. (2007, July). Virtual reality: How much immersion is enough? *Computer*, *40*(7), 36–43.

Retrieved from `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4287241` doi: 10.1109/MC.2007.257

Bricken, W. (1989). Virtual reality: Directions of growth. In *Notes from the siggraph* (pp. 1–13).

Brown, R. F., & Bylund, C. L. (2008). Communication skills training: Describing a new conceptual model. *Academi Medicine*, *83*(1), 37–44.

Burdea, G. C., & Coiffet, P. (2003). Virtual reality: Technology. In *Virtual reality: Technology* (2nd ed., pp. 285–348). Hoboken, NJ: John Wiley & Sons.

Bush, J. (2008, May). Viability of virtual reality exposure therapy as a treatment alternative. *Computers in Human Behavior*, *24*(3), 1032–1040. Retrieved from `http://linkinghub.elsevier.com/retrieve/pii/S0747563207000787` doi: 10.1016/j.chb.2007.03.006

Butler, J. L., & Baumeister, R. F. (1998). The Trouble with friendly faces: Skilled performance with a supportive audience. *Journal of Personality and Social Psychology*, *75*(5), 1213–1230. doi: 10.1037/0022-3514.75.5.1213

Cassell, J. (2001). Embodied conversational agents: Representation and intelligence in user interfaces. *AI Magazine*, *22*(4), 67–84. Retrieved from `http://www.justinecassell.com/publications/AIMag22-04-007.PDF`

Cassell, J., Sullivan, J., M., P., & Churchill, E. (2000). *Embodied conversational agents* (Authors, Ed.). Cambridge, MA: The MIT Press.

Cassell, J., & Vilhjálmsson, H. H. (1999). Fully embodied conversational avatars: Making communicative behaviors autonomous. *Autonomous Agents and Multi-Agent Systems*, 45–64. Retrieved from `http://link.springer.com/article/10.1023/A:1010027123541` doi: 10.1023/A:1010027123541

Chant, S., Tim, Randle, J., Russell, G., & Webb, C. (2002, April). Communication skills training in healthcare: A Review of the literature. *Nurse Education Today*, *22*(3), 189–202. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/12027600` doi: 10.1054/nedt.2001.0690

Charmaz, K. (1996). The search of meanings: Grounded theory. In S. R. Harré & L. Van Langenhove (Eds.), *Rethinking methods in psychology* (pp. 27–49). London: Sage. doi: 10.1016/B978-0-08-044894-7.01581-5

Chartrand, T. L., & Bargh, J. A. (1999). The Chameleon effect: The Perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893–910. Retrieved from `http://www.yale.edu/acmelab/articles/chartrand{_}bargh{_}1999.pdf` doi: 10.1037/0022-3514.76.6.893

Christophel, D. M. (1990). The Relationships among teacher immediacy behaviors, student motivation, and learning. *Communication Education*, *39*(4), 323–340. doi: 10.1080/03634529009378813

Chun, M. M., & Wolfe, J. M. (2001). Visual attention. In B. Goldstein (Ed.), *Blackwell handbook of perception* (pp. 272–310). Oxford: Blackwell Publishers Ltd.

Cialdini, R. B., Borden, R. J., Thorne, A., Walker, M. R., Freeman, S., & Sloan, L. R. (1976). Basking in reflected glory: Three (football) field studies. *Journal of Personality and Social Psychology*, *34*, 366–375. doi: 10.1037/0022-3514.34.3.366

Clayman, S. E. (1993). Booing: The Anatomy of a disaffiliative response. *American Sociological Review*, *58*(1), 110–130. Retrieved from `http://www.jstor.org.proxy.lib.sfu.ca/stable/2096221`

Clevenger Jr., T., King, T. R., Clevenger, T., & King, T. R. (1961). A Factor analysis of the visible symptoms of stage fright. *Speech Monographs*, *28*(4), 296–298. doi: 10.1080/03637756109375328

Cohen, J. (1960). A Coefficient of agreement for nominal scales. *Education and Psychological Measurement*, *XX*(1), 37–46. doi: 10.1177/001316446002000104

Cohen, J. L., & Davis, J. (1973). Effects of audience status, evaluation, and time of action on performance whtih hidden-word problems. *Journal of Personality and Social Psychology*, *27*(1), 74–85. Retrieved from `http://dx.doi.org/10.1037/h0034447`

Cohn, J., Reed, L., Moriyama, T., Xiao, J., Schmidt, K., & Ambadar, Z. (2004). Multimodal coordination of facial action, head rotation, and eye motion during spontaneous smiles. *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.*, 129–135. Retrieved from `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1301520` doi: 10.1109/AFGR.2004.1301520

Compton, D., Love, T. P., & Sell, J. (2012, November). Developing and assessing intercoder reliability in studies of group

interaction. *Sociological Methodology*, *42*(1), 348–364. Retrieved from `http://smx.sagepub.com/lookup/doi/10.1177/0081175012444860` doi: 10.1177/0081175012444860

Cook, D. a., & Triola, M. M. (2009). Virtual patients: A Critical literature review and proposed next steps. *Medical Education*, *43*(4), 303–311. doi: 10.1111/j.1365-2923.2008.03286.x

Coulson, M. (2004). Attributing emotion to static body postures: Recongition accuracy, confusion and view point dependence. *Journal of Nonverbal Behavior*, *28*(2), 117–139.

Cronin, M., & Glenn, P. (1991). Oral communication across the curriculum in higher education: The State of the art. *Communication Education*, *40*(October), 356–367.

Cruz-Neira, C., Sandin, D. J., & Defanti, T. A. (1993). Surround-screen projection-based virtual reality: The Design and implementation of the CAVE. In *SIGGRAPH '93 Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 135–142). New York. doi: 10.1145/166117.166134

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco: Jossey-Bass Publishers.

Cummings, J. J., Bailenson, J. N., & Fidler, M. J. (2015). How immersive is enough? A Meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, 1–57. doi: 10.1.1.363.6971

Dael, N., Mortillaro, M., & Scherer, K. R. (2012, January). The Body Action and Posture coding system (BAP): Development and reliability. *Journal of Nonverbal Behavior*, *36*(2), 97–121. Retrieved from `http://link.springer.com/10.1007/s10919-012-0130-0` doi: 10.1007/s10919-012-0130-0

Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology*, *41*(1), 10–32. doi: 10.1111/j.1467-8535.2009.01038.x

Daly, J. A., Vangelisti, A. L., & Lawrence, S. G. (1989, January). Self-focused attention and public speaking anxiety. *Personality and Individual Differences*, *10*(8), 903–913. Retrieved from `http://linkinghub.elsevier.com/retrieve/pii/0191886989900251` doi: 10.1016/0191-8869(89)90025-1

Dance, F. E. X. (1987, November). What do you mean presentational skills? *Management Communication Quarterly*, *1*(2), 260–271. Retrieved from `http://mcq.sagepub.com/cgi/doi/10.1177/`

0893318987001002006 doi: 10.1177/0893318987001002006

Dannels, D. P. (2001, April). Time to speak up: A Theoretical framework of situated pedagogy and practice for communication across the curriculum. *Communication Education*, *50*(2), 144–158. doi: 10.1080/03634520109379240

Dannels, D. P. (2002, July). Communication across the curriculum and in the disciplines: speaking in engineering. *Communication Education*, *51*(3), 254–268. doi: 10.1080/03634520216513

Dannels, D. P., & Housley Gaffney, A. L. (2009). Communication across the curriculum and in the disciplines: A Call for scholarly cross-curricular advocacy. *Communication Education*, *58*(1), 124–153. doi: 10.1080/03634520802527288

Darling, A. L., & Dannels, D. P. (2003). Practicing engineers talk about the importance of talk: A Report on the role of oral communication in the workplace. *Communication Education*, *52*(1), 1–16. doi: 10.1080/03634520302457

Darwin, C. (2000). *Der Ausdruck der Gemütsbewegungen bei dem Menschen und den Tieren* (7th ed.; P. Ekman, Ed.). Frankfurt a.M.: Eichborn.

Deacon, B., & Abramowitz, J. (2006). A Pilot study of two-day cognitive-behavioral therapy for panic disorder. *Behaviour Research and Therapy*, *44*(6), 807–817. doi: 10.1016/j.brat.2005.05.008

Delia, J. (2003). Constructivism. In H. Griffin (Ed.), *A first look at communication theory* (Eight ed., pp. 98–112). McGraw-Hill.

Derrick, D. C., & Ligon, G. S. (2014). The Affective outcomes of using influence tactics in embodied conversational agents. *Computers in Human Behavior*, *33*, 39–48. Retrieved from http://dx.doi.org/10.1016/j.chb.2013.12.027 doi: 10.1016/j.chb.2013.12.027

de Vaus, D. (2001). What is research design. In *Research design in social reseach.* London: Sage. doi: 10.1177/1473325011401471

Di Salvo, V. S. (1980). A Summary of current research identifying communication skills in various organizational contexts. *Communication Education*. doi: 10.1080/03634528009378426

Difede, J., & Hoffman, H. G. (2002, December). Virtual reality exposure therapy for World Trade Center post-traumatic stress disorder: A Case report. *Cyberpsychology & Behavior*, *5*(6), 529–35. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12556115

Docan-Morgan, T. (2009). I Now see how I can use these skills: An Applied project for the public speaking course. *Communication Teacher*, *23*(3), 110–116. doi: 10.1080/17404620902974808

Dunbar, N. E., Brooks, C. F., & Kubicka-Miller, T. (2006, July). Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills. *Innovative Higher Education*, *31*(2), 115–128. Retrieved from `http://link .springer.com/10.1007/s10755-006-9012-x`   doi: 10.1007/ s10755-006-9012-x

Duncan, S. (1974). On the structure of speaker-auditor interaction during. *Language in Society*, *3*(2), 161–180.

Duncan Jr., S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, *23*(2), 283–292.

Dunston, P. S., Arns, L., & McGlothlin, J. D. (2007). An Immersive virtual reality mock-up for design review of hospital patient rooms. In *7th International Conference on Construction Applications of Virtual Reality* (pp. 1–9). Retrieved from `https://www.engr.psu.edu/convr/proceedings/ papers/01{_}Dunston{_}submission{_}45.pdf`

Dwyer, K. K., & Davidson, M. M. (2012, April). Is public speaking really more feared than death? *Communication Research Reports*, *29*(2), 99–107. doi: 10.1080/08824096.2012.667772

Eastgate, R. M., Wilson, J. R., & D'Cruz, M. (2014). Structured development of virtual environments. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (Second ed., pp. 353–389). CRC Press.

Eimer, M., Nattkemper, D., Schröger, E., & Prinz, W. (1996). Involuntary attention. In O. Neumann & A. F. Sanders (Eds.), *Handbook of perception and action* (pp. 155–184).

Ekman, P. (1979). About brows: Emotional and conversational signals. In M. von Cranah, K. Foppa, W. Lepenies, & D. Ploog (Eds.), *Human ethology* (pp. 169–207). Cambridge: University Press. doi: 10.2307/40538147?ref=no-x-route: d64912790719373b34a8db686a88913d

Ekman, P., & Friesen, W. (1969). The Repertoire of nonverbal behavior, categories, origins, usage, and coding. *Semiotica*, *1*(1), 49–98. Retrieved from `http://www.paulekman.com/wp-content/ uploads/2009/02/The-Repertoire-Of-Nonverbal-Behavior -Categories-Origins-.pdf`

Elkins, A. C., Derrick, D. C., Burgoon, J. K., & Nunamaker Jr., J. F. (2012, jan). Predicting users' perceived trust in embodied conversational agents using vocal dynamics. *2012 45th Hawaii International Conference on System Sciences*, 579–588. doi: 10.1109/HICSS.2012.483

Emmelkamp, P. M. G., Krijn, M., Hulsbosch, a. M., De Vries, S. D., Schuemie, M. J., & van der Mast, C. A. P. G. (2002). Virtual reality treatment versus exposure in vivo: A Comparative evaluation in acrophobia. *Behaviour Research and Therapy*, *40*(5), 509–516. doi: 10.1016/S0005-7967(01)00023-7

Endsley, M. R., & Jones, D. G. (2011). User-centered design. In C. Press (Ed.), *Designing for situation awareness an approach to user-centered design* (Second ed., pp. 3–12).

Fabri, M., Moore, D. J., & Hobbs, D. J. (1999). The Emotional avatar: Non-verbal communication between inhabitants of collaborative virtual environments. In A. Braffort, R. Gherbi, S. Gibet, D. Teil, & J. Richardson (Eds.), *International gesture workshop* (pp. 269–273). Gif-sur-Yvette: Springer-Verlag. doi: 10.1007/3-540-46616-9{_}24

Faravelli, C., Zucchi, T., Viviani, B., Salmoria, R., Perone, A., & Paionni, A. (2000). Epidemiology of social phobia: A Clinical approach. *European Psychiatry*, 17–24. doi: 10.1016/S0924-9338(00)00215-7

Farkas, I., Helbing, D., & Vicsek, T. (2002, September). Mexican waves in an excitable medium. *Nature*, *419*(6903), 131–2. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12226653 doi: 10.1038/419131a

FASP. (2014). *Kompetenzzentrum Virtual Reality Ilmenau.* Retrieved from (http://www.tu-ilmenau.de/kvr/technische-ausstattung/

Faulknera, A., Webbb, P., & Maguire, P. (1991). Communication and counseling skills: Educating health professionals working in cancer and palliative care. *Patient Education and Counseling*, *18*, 3–7. doi: 10.1016/0738-3991(91)90073-E

Forsyth, D. R. (2005). *Group Dynamics* (4th ed.). Wadsworth Publishing.

Forsyth, D. R. (2009). *Group Dynamics* (5th ed.). Belmont: Wadsworth Publishing.

Fremouw, W. J., & Zitter, R. E. (1978). A Comparison of skills training and cognitive restructuring-relaxation for the treatment

of speech anxiety. *Behavior Therapy*, *9*(2), 248–259. doi: 10.1016/S0005-7894(78)80110-5

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007, July). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*(4), 694–724. doi: 10.1037/0033-2909.133.4.694

Fuchs, P., Moreau, G., & Guitton, P. (2011). *Virtual reality: Concepts and technologies*. CRC Press.

Furbay, A. L. (1965). The Influence of scattered versus compact seating on audience response. *Speech Monographs*, *32*(2), 144–148.

Gabbard, J. L., Hix, D., & Swan II, J. E. (1999). User-centered design and evaluation of virtual environments. *Computer Graphics and Applications, IEEE*, *19*(6), 51–59. doi: 10.1109/38.799740

Garcia-Leal, C., Parente, A. C. B. V., Del-Ben, C. M., Guimaraes, F. S., Moreira, A. C., Elias, L. L. K., & Graeff, F. G. (2005, feb). Anxiety and salivary cortisol in symptomatic and nonsymptomatic panic patients and healthy volunteers performing simulated public speaking. *Psychiatry Research*, *133*(2-3), 239–52. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/15740999` doi: 10.1016/j.psychres.2004.04.010

Garcia-Palacios, A., Hoffman, H., Carlin, A., Furness III, T. A., & Botella, C. (2002). Virtual reality in the treatment of spider phobia: A Controlled study. *Behavior Research and Therapy*, *40*, 983–993. doi: 10.1016/S0005-7967(01)00068-7

Gaudiosi, J. (2015). Virtual reality now actual reality with four vr headset launches coming soon. *Fortune*. Retrieved from `http://fortune.com/2015/05/07/vr-headset-launches/`

Gerrig, R. J., & Zimbardo, P. G. (2002). *Glossary of psychological terms* (16th ed.). Boston: Allyn & Bacon. Retrieved from `http://www.apa.org/research/action/glossary.aspx`

Gifford, R. (1994a). A Lens-mapping framework for understanding the encoding and decoding of interpersonal dispositions in nonverbal behavior. *Journal of Personality and Social Psychology*, *66*(2), 398–412. doi: 10.1037/0022-3514.66.2.398

Gifford, R. (1994b). *SKANS V: Seated kinesics activity notation system.*

Glanz, K., Rizzo, A. S., & Graap, K. (2003). Virtual reality for psychotherapy: Current reality and future responsibilities. *Psychotherapy: Theory/Research/Practice/*, *40*(1/2), 1–13. Retrieved from `http://032912b.membershipsoftware.org/`

libdocuments/Virtual{_}Reality{_}Uses.pdf

Gläser, J., & Laudel, G. (2013). Life with and without coding: Two methods for early-stage data analysis in qualitative research aiming at casual explanations. *Forum Qualitative Sozialforschung / Forum Qualitative Social Research*, *14*(2). Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/1886/3528{#}g312

Google. (2016). *Google Developers.* Retrieved 2001-06-20, from https://developers.google.com/cardboard/?hl=en

Gratch, J., & Marsella, S. (2004). Evaluating the modeling and use of emotion in virtual humans. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1* (p. 327). doi: 10.1109/AAMAS.2004.242331

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. doi: 10.1109/MIS.2002.1024753

Green, J. O. (2003). Models of adult communication skill acquisition: Practice and the course of performance improvement. In J. O. Greene & B. R. Burleson (Eds.), *Handbook of communication and social interaction skills.* Mahwah, N.J.: Routledge.

Grillon, H., Riquier, F., Herbelin, B., & Thalmann, D. (2006). Use of virtual reality as therapeutic tool for behavioural exposure in the ambit of social anxiety disorder treatment. In *6th intl. conf. disability, virtual reality and assoc. tech.* (pp. 105–112). Esbjerg.

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. Retrieved from http://www.eyethink.org/resources/papers/Gueguen-et-al..pdf

Gulliksen, J., Göransson, B., Boivie, I., Blomkvist, S., Persson, J., & Cajander, Å. (2003). Key principles for user-centred systems design. *Behaviour & Information Technology*, *22*(6), 397–409. doi: 10.1080/01449290310001624329

Gutierrez, D., Frischer, B., Cerezo, E., Gomez, A., & Seron, F. (2007). AI and virtual crowds: Populating the Colosseum. *Journal of Cultural Heritage*, *8*(2). doi: 10.1016/j.culher.2007.01.007

Hale, J. L., & Burgoon, J. K. (1984). Models of reactions to changes in nonverbal immediacy. , *8*(4), 287–314.

Hargie, O. (Ed.). (2006). *The Handbook of communication skills* (3rd ed.). London: Routledge.

Harrigan, J. a., Rosenthal, R., & Scherer, K. R. (Eds.). (2008). *The New handbook of methods in nonverbal behavior research.* doi: 10.1093/acprof:oso/9780198529620.001.0001

Harris, S. R., Kemmerling, R. L., & North, M. (2002, December). Brief virtual reality therapy for public speaking anxiety. *Cyberpsychology & Behavior*, *5*(6), 543–50. doi: 10.1089/109493102321018187

Harrow, T. L., & Dziuban, C. D. (1974). An Investigation into the relationship of student teacher's perceived problems to those of supervising teachers. In *Annual Meeting of the American Educational Research Association.* Chicago.

Hartmann, B., Mancini, M., & Pelachaud, C. (2005). Implementing expressive gesture synthesis for. In S. Gibet, N. Courty, & J.-F. Kamp (Eds.), *Gesture in human-computer interaction and simulation 6th international gesture workshop* (pp. 188–199). Berder Island: Springer-Verlag. Retrieved from `http://www.springerlink.com/content/978-3-540-32624-3/{#}section=498892{&}page=1`

Hartmann, B., Mancini, M., & Pelachaud, C. (2006). Implementing expressive gesture synthesis for embodied conversational agents. *Lecture Notes in Artificial Intelligence*, *3881*, 188–199. doi: 10.1007/11678816{_}22

Heimberg, R. G. (2002, January). Cognitive-behavioral therapy for social anxiety disorder: Current status and future directions. *Biological Psychiatry*, *51*(1), 101–8. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/11801235`

Heimberg, R. G., Holt, C. S., Schneier, F. R., Spitzer, R. L., & Liebowitz, M. R. (1993). The Issue of subtypes in the diagnosis social phobia. *Journal of Anxiety Disorders*, *7*, 249–269.

Heimberg, R. G., Hope, D. A., Dodge, C. . S., & Becker, R. E. (1990). DSM-III-R subtypes of social phobia. Comparison of generalized social phobics and public speaking phobics. *The Journal of Nervous and Mental Disease*, *178*(3), 172–9.

Helander, M. (2006). *A Guide to human factors and ergonomics* (second ed.). Boca Raton: Taylor & Francis. doi: 10.1080/00140130701680379

Hennig, J., Netter, P., & Voigt, K.-H. (2001, oct). Cortisol mediates redistribution of CD8+ but not of CD56+ cells after the psychological stress of public speaking. *Psychoneuroendocrinology*, *26*(7), 673–687. Retrieved from `http://linkinghub.elsevier`

.com/retrieve/pii/S0306453001000208 doi: 10.1016/S0306
-4530(01)00020-8

Heritage, J., & Greatbatch, D. (1986). Generating applause: A study of rhetoric and response at party political conferences. *AJS*, *92*(1), 110–157.

Hess, U., Blairy, S., & Kleck, R. E. (1997). The Intensity of emotional facial expressions and decoding acccuracy. *Journal of Nonverbal Behavior*, *21*(4), 241–257.

Heylen, D. (2009). Understanding speaker-listener interactions. In *Interspeech* (pp. 2151–2154). Brighton.

Hilmert, C. J., Christenfeld, N., & Kulik, J. A. (2002). Audience status moderates the effects of social support and self-efficacy on cardiovascular reactivity during public speaking. *The Society of Behavioral Medicine*, 122–131.

Hocking, J. E., Margreiter, D. G., & Hylton, C. (1977). *Intra-audiece effects: A Field test* (Vol. 3) (No. 3).

Hoffman, M., & Mittelman, M. (2004, January). Presentations at professional meetings: Notes, suggestions and tips for speakers. *European Journal of Internal Medicine*, *15*(6), 358–363. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15522569 doi: 10.1016/j.ejim.2004.07.007

Hoffman, R. R. (1987). The Problem of extracting the knowledge of experts from the perspective of experimental psychology. , *8*(2), 53–67.

Hoffman, R. R. (1989). A Survey of methods for eliciting the knowledge of experts. *SIGART Newsletter*(108), 19–27.

Huang, H.-H. (2009). *A Generic framework for embodied conversational agent development and its applications.*

Hughes, I., & Large, B. (1993, January). Staff and peer-group assessment of oral communication skills. *Studies in Higher Education*, *18*(3), 379–385. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/03075079312331382281 doi: 10.1080/03075079312331382281

Hylton, C. A. L. (1971). Intra-audience effects: Observable audience response. *The Journal of Communication*, *21*(September), 253–265.

Ijsselsteijn, W., & Riva, G. (2003). Being there: The Experience of presence in mediated environments. In G. Riva, F. Davide, & W. A. Ijsselsteijn (Eds.), *Being there: Concepts, effects and measurement of user presence in synthetic environments* (p. 14).

Amsterdam: IOS Press.

Jackson, J. M., & Latane, B. (1981). All alone in front of all those people: Stage fright as a function of number and type of co-performers and audience. *Journal of Personality and Social Psychology*, *40*(1), 73–85. doi: 10.1037/0022-3514.40.1.73

Jantke, K. P., Knauf, R., & Gonzalez, A. J. (2006). Storyboarding for playful learning. In T. C. Reeves & S. F. Yamashita (Eds.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2006 (E-Learn 2006)* (pp. 3174–3182). Honolulu.

Jerald, J. (2015). *The VR Book: Human-Centered Design for Virtual Reality*. Retrieved from http://dl.acm.org/citation.cfm?id=2792790

Johnsen, K., Dickerson, R., Raij, a., Lok, B., & Jackson, J. (2005). Experiences in using immersive virtual characters to educate medical communication skills. *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, *2005*, 179–187. doi: 10.1109/VR.2005.1492772

Johnson, Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, *11*(July 1999), 47–78. Retrieved from http://ijaied.org/pub/946/file/946{_}paper.pdf

Johnson, & Szczupakiewicz, N. (1987). The Public speaking course: Is it preparing students with work related public speaking skills? *Communication Education*. doi: 10.1080/03634528709378653

Jonassen, D. (1998). Designing constructivist learning environments. In C. Reigeluth (Ed.), *Instructional-design theories and models* (Second ed., pp. 215 – 239). Mahwah, NJ: Lawrence Erlbaum.

Kang, N., Brinkman, W.-P., van Riemsdijk, M. B., & Neerincx, M. A. (2013, oct). An Expressive virtual audience with flexible behavioral styles. *IEEE Transactions on Affective Computing*, *4*(4), 326–340. Retrieved from http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6714373 doi: 10.1109/TAFFC.2013.2297104

Kasap, Z., & Magnenat-Thalmann, N. (2007). Intelligent virtual humans with autonomy and personality: State-of-the-art. *Studies in Computational Intelligence*, *1*, 3–15. doi: 10.1007/978-3-540-79868-2{_}2

Kaur, K. (1998). *Designing virtual environments for usability*

(Unpublished doctoral dissertation). City University London.

Kawulic, B. B. (2005). Participant observation as a data collection method. *Forum: Qualitative social research*, *6*(2). Retrieved from `http://www.qualitative-research.net/index.php/fqs/article/view/466/996`

Kenny, Parsons, T. D., Gratch, J., Leuski, A., & Rizzo, A. A. (2007). Virtual patients for clinical therapist skills training. In *IVA 2007* (pp. 197–210). doi: 10.1007/978-3-540-74997-4{_}19

Kenny, Parsons, T. D., & Rizzo, A. (2009). A comparative analysis between experts and novices interacting with a virtual patient with PTSD. *Annual Review of CyberTherapy and Telemedicine*, *7*(1), 122–124. doi: 10.3233/978-1-60750-017-9-122

Kenny, P., Hartholt, A., Gratch, J., Swartout, W., Traum, D., Marsella, S., & Piepol, D. (2007). Building interactive virtual humans for training environments. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)* (pp. 1–16).

Kessler, R. C., Stein, M. B., & Berglund, P. (1998). Social phobia subtypes in the national comorbidity survey. *American Journal of Psychiatry*, *155*(5), 613–619.

Kipp, M. (2003). *Gesture generation by imitation: From human behavior to computer character animation* (Unpublished doctoral dissertation).

Kiselev, A., Hacker, B. A., Wankerl, T., Abdikeev, N., & Nishida, T. (2010, November). Toward incorporating emotions with rationality into a communicative virtual agent. *AI & Society*. Retrieved from `http://www.springerlink.com/index/10.1007/s00146-010-0309-z` doi: 10.1007/s00146-010-0309-z

Knapp, M., & Hall, J. (2010). *Nonverbal communication in human interaction* (7th ed.). Boston, MA: Wadsworth, Cengage Learning.

Knowles, E. S., & Bassett, R. L. (1976). Groups and crowds as social entities: Effects of activity, size, and member similarity on nonmembers. *Journal of Personality and Social Psychology*, *34*(5), 837–845. doi: 10.1037//0022-3514.34.5.837

Koholka, C. (2012). *Angstfreies Erzählen im tiergestützten Deutschunterricht: Didaktische und methodische Ausführungen.*

Kothgassner, O. D., Felnhofer, A., Beutl, L., Hlavacs, H., Lehenbauer, M., & Stetina, B. (2012). A Virtual training tool for giving talks. *Lecture Notes in Computer Science*, *7522 LNCS*, 53–66. doi: 10.1007/978-3-642-33542-6{_}5

Kowal, S. (2009). *Communicating with one another: Toward a*

psychology of spontaneous spoken discourse. Springer.

Krauss, R. M., Chen, Y., & Chawla, P. (2004). Nonverbal behavior and nonverbal communication: What do conversations hand gestures tell us? In Zanna (Ed.), *Advances in experimental social psychology* (pp. 389–450). San Diego, CA: Academic Press.

Krijn, M., Emmelkamp, P., Olafsson, R., & Biemond, R. (2004). Virtual reality exposure therapy of anxiety disorders: A Review. *Clinical Psychology Review*, *24*, 259 – 281. doi: 10.1016/j.cpr.2004.04.001

Krijn, M., Emmelkamp, P. M. G., Biemond, R., de Wilde de Ligny, C., Schuemie, M. J., & van der Mast, C. A. P. G. (2004, feb). Treatment of acrophobia in virtual reality: The Role of immersion and presence. *Behaviour research and therapy*, *42*(2), 229–39. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/14975783` doi: 10.1016/S0005-7967(03)00139-6

Krishnan, V., Foster, A., Kopper, R., & Lok, B. (2012). Virtual human personality masks: A Human computation approach to modeling verbal personalities in virtual humans. In *Intelligent Virtual Agents. 12th International Conference, IVa 2012, santa cruz, ca, usa, september, 12-14, 2012.* (pp. 146–152).

Kshirsagar, S., & Magnenat-Thalmann, N. (2002). A Multilayer personality model. In *SMARTGRAPH '02 Proceedings of the 2nd International Symposium on Smart Graphics* (pp. 107–115). doi: 10.1145/569005.569021

Lambertz, K. (2011). *Back channelling: The use of yeah and mm to portray engaged listenership* (Vol. 2). Retrieved from `https://www.google.de/url?sa=t{&}rct=j{&}q={&}esrc=s{&}source=web{&}cd=1{&}cad=rja{&}uact=8{&}ved=0CCIQFjAA{&}url=http://www.griffith.edu.au/{_}{_}data/assets/pdf{_}file/0005/384017/Lambertz-backchannelling.pdf{&}ei=9wptVZvlPIzTU{_}C0gbAG{&}usg=AFQjCNEeaoUQXYlXCLDhM9c6Qj8GE6{_}hAw{&}sig2=`

Lanier, J. (1989). An Interview with Jaron lanier. *Whole Earth Review*.

Latane, B., & Harkins, S. (1976). Cross-modality matches suggest anticipated stage fright a multiplicative power function. *Perception & Psychophysics*, *20*(6), 482–488.

Latane, B., & Zipf, G. K. (1981). The Psychology of social impact. *American Psychologist*, *36*(4), 343–356. doi: http://dx.doi.org/10.1037/0003-066X.36.4.343

Lawrence, J. B. (1979). Smiles can be back channels. *Journal of Personality and Social Psychology*, *37*(5), 728–734. doi: 10.1037/0022-3514.37.5.728

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 & Behaviour*, *5*(3), 191–195. Retrieved from `www.ncbi.nlm.nih.gov/pubmed/12123239`

Lee, K. M. (2004). Presence, Explicated. *Communication Theory*, *14*(1), 27–50. Retrieved from `http://onlinelibrary.wiley.com/doi/10.1111/j.1468-2885.2004.tb00302.x/epdf` doi: 10.1093/ct/14.1.27

Levasseur, D. G., Dean, K. W., & Pfaff, J. (2004, July). Speech pedagogy beyond the basics: A Study of instructional methods in the advanced public speaking course. *Communication Education*, *53*(3), 234–252. doi: 10.1080/0363452042000265170

Loftin, R. B., Scerbo, M. W., McKenzie, R., Catanzaro, J. M., Bailey, N. R., Phillips, M. A., & Perry, G. (2003). Training in peacekeeping operations using virtual environments. In *RTO HFM Symposium on Advanced Technologies for Military Training.* Genoa. Retrieved from `http://ftp.rta.nato.int/public/PubFullText/RTO/MP/RTO-MP-HFM-101/MP-HFM-101-03.pdf`

Lombard, & Ditton. (1997). At the heart of it all. *Journal of Computer-Mediated Communication*, *3*(2). doi: 10.1111/j.1083-6101.1997.tb00072.x

Lourens, T., van Berkel, R., & Barakova, E. (2010, December). Communicating emotions and mental states to robots in a real time parallel framework using Laban movement analysis. *Robotics and Autonomous Systems*, *58*(12), 1256–1265. Retrieved from `http://linkinghub.elsevier.com/retrieve/pii/S092188901000151X` doi: 10.1016/j.robot.2010.08.006

Lucas, S. E. (2012). *The Art of public speaking* (11th ed.). McGrall Hill.

MacLean, K. a., Aichele, S. R., Bridwell, D. a., Mangun, G. R., Wojciulik, E., & Saron, C. D. (2009). Interactions between endogenous and exogenous attention during vigilance. *Attention, Perception and Psychophysics*, *71*(5), 1042–1058. doi: 10.3758/APP.71.5.1042

Maclntyre, P. D., & Thivierge, K. A. (1995). The Effects of audience pleasantness, audience familiarity, and speaking contexts on

public speaking anxiety and willingness to speak. *Communication Quarterly*, *43*(4), 456–466. doi: 10.1080/01463379509369992

Magnenat-Thalmann, N., & Thalmann, D. (2004). *Handbook of virtual humans* (N. Magnenat-Thalmann & D. Thalmann, Eds.). Chichester: John Wiley and Sons.

Maldonado, H., & Hayes-Roth, B. (2004). Toward cross-cultural believability in character design. *Agent Culture: Human-Agent Interaction in a Multicultural World*, 143–176.

Manganas, A., Tsiknakis, M., Leisch, E., Ponder, M., Molet, T., Herbelin, B., . . . Schenone, A. (2004). The JUST VR tool: An Innovative approach to training the citizen for emergency situations using virtual reality techniques. *The Journal on Information Technology in Healthcare*, *2*(6), 399–412. Retrieved from `http://www.forth.gr/ics/bmi/publications/ papers/2004/ICMCC{_}JUST/ICMCC{_}JUSTWebCD{_}CR .pdf$\delimiter"026E30F$npapers3://publication/uuid/ D4DF9AEC-1664-4399-9B2D-EBD2EFC18C85`

Mann, S., & Robinson, A. (2009). Boredom in the lecture theatre: An Investigation into the contributors, moderators and outcomes of boredom amongs university students. *British Educational Research Journal*, *35*(2), 243–258.

Mantovani, F., & Castelnuovo, G. (2003). *The Sense of presence in virtual training: Enhancing skills acquisition and transfer of knowledge through learning experience in virtual environments.* Amsterdam: IOS Press. Retrieved from `http://publicatt .unicatt.it/handle/10807/32450`

Mantovani, G., & Riva, G. (1999). Real presence: How different ontologies generate different criteria for presence, telepresence, and virtual presence. *Presence: Teleoperators and Virtual Environments*, *8*(5), 540–550. doi: 10.1162/105474699566459

Martin, J.-C., Niewiadomsky, R., Devillers, L., Buisine, S., & Pelachaud, C. (2006). Multimodal complex emotions: Gesture expressivity and blended facial expressions. *International Journal of Humanoid Robotics*, *3*(3), 269–291. Retrieved from `http://citeseerx.ist.psu.edu/viewdoc/ download?doi=10.1.1.95.7260{&}rep=rep1{&}type=pdf`

McCroskey, J. (1970). Measures of communication-bound anxiety. *Speech Monographs*, *37*, 269–277. Retrieved from `http://www .jamescmccroskey.com/measures/prpsa.htm`

McCroskey, J. (1977). Oral commuication apprehension: A Summary

of recent theory and research. *Human Communication Research*, *4*(1), 78–96. Retrieved from `http://dx.doi.org/10.1111/j.1468-2958.1977.tb00599.x` doi: 10.1111/j.1468-2958.1977.tb00599.x

McCroskey, J., & Daly, J. (1984). The Communication apprehension perspective. In J. A. Daly & J. C. McCroskey (Eds.), *Avoiding communications: Shyness, reticence, and communication apprehension* (pp. 13–38). Beverly Hills: Sage. Retrieved from `https://www.ffri.hr/{~}ibrdar/komunikacija/seminari/McCroskey,1984-Communicationapprehensionperspective(Ch.pdf`

McKinney, M. E., Gatchel, R. J., & Paulus, B. (1983). The Effects of audience size on high and low speech-anxious subjects during an actual speaking task. *Basic and Applied Social Psychology*, *4*(1), 73–87.

McMahan, R. P., Bowman, D. A., Zielinski, D. J., & Brady, R. B. (2012, April). Evaluating display fidelity and interaction fidelity in a virtual reality game. *IEEE Transactions on Visualization and Computer Graphics*, *18*(4), 626–33. doi: 10.1109/TVCG.2012.43

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* Chicago: The University of Chicago Press.

McPhail, C. (2014). Blumer's theory of collective behavior: The Development of a non-symbolic interaction explanation. *The Sociological Quarterly*, *30*(3), 401–423. Retrieved from `http://www.jstor.org/stable/4120849`

McPhail, C., Powers, W. T., & Tucker, C. W. (1992, April). Simulating individual and collective action in temporary gatherings. *Social Science Computer Review*, *10*(1), 1–28. doi: 10.1177/089443939201000101

Meyerbröker, K., Morina, N., Kerkhof, G., & Emmelkamp, P. M. G. (2011). Virtual reality exposure treatment of agoraphobia: A Comparison of computer automatic virtual environment and head-mounted display. , 51–57. doi: 10.3233/978-1-60750-766-6-51

Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A Ten-year review of empirical research (1999-2009). *Computers and Education*, *56*(3), 769–780. Retrieved from `http://dx.doi.org/10.1016/j.compedu.2010.10.020` doi: 10.1016/j.compedu.2010.10.020

Mikropoulos, T. A., & Strouboulis, V. (2004). Factors that influence presence in educational virtual environments. *Cyberpsychology & Behavior*, *7*(5), 582–591. doi: 10.1089/cpb.2004.7.582

Miles, L., & Johnston, L. (2007, sep). Detecting happiness: Perceiver sensitivity to enjoyment and non-enjoyment smiles. *Journal of Nonverbal Behavior*, *31*(4), 259–275. doi: 10.1007/s10919-007 -0036-4

Miller, T. C., & Stone, D. N. (2009). Public speaking apprehension (PSA), motivation, and affect among accounting majors: A Proof-of-concept intervention. *Issues in Accounting Education*, *24*(3), 265–298. doi: 10.2308/iace.2009.24.3.265

Mitchell, B. R. (2012). *Game design essentials*. Sybex.

Mori, M. (1970). The Uncanny valley. *Energy*, *7*(4), 33 – 35. Retrieved from `http://www.androidscience.com/ theuncannyvalley/proceedings2005/uncannyvalley.html`

Morreale, S. P., & Backlund, P. M. (2002, January). Communication curricula: History, recommendations, resources. *Communication Education*, *51*(1), 2–18. doi: 10.1080/03634520216498

Morreale, S. P., Hanna, M. S., Berko, R. M., & Gibson, J. W. (1999). The Basic communication course at U.S. colleges and universities: VI. *Basic Communication Course Annual*, *11*, 1–36.

Morreale, S. P., Moore, M., Surges-Tatum, D., & Webster, L. (Eds.). (2007). *The Competent speaker speech evaluation form.*

Morreale, S. P., Rubin, R. B., & Jones, E. (1998). *Speaking and listening competencies for college students.*

Moulin, B., & Brassard, M. (1996). A Scenario-based design method and an environment for the development of multiagent systems. *Distributed Artificial Intelligence Architecture and Modelling*(Lecture Notes in Computer Science 1087), 216–232. Retrieved from `http://link.springer.com/chapter/10.1007/ 3-540-61314-5{_}32` doi: 10.1007/3-540-61314-5_32

Myers, S. a., & Ferry, M. F. (2001, March). Interpersonal communication motives and nonverbal immediacy behaviors. *Communication Research Reports*, *18*(2), 182–191. doi: 10.1080/ 08824090109384796

Nooraei, B., Rich, C., & Sidner, C. L. (2014). A real-time architecture for embodied conversational agents: Beyond turn-taking. In *Proceedings of the 7th international conference on advances in computer-human interactions (achi).*

North, M. M., North, S. M., & Coble, J. R. (1998). Virtual reality

therapy: An Effective treatment for the fear of public speaking. *The International Journal of Virtual Reality*, 1–6.

Norton, R. W., & Pettegrew, L. S. (1979). Attentiveness as a style of communication: A Structural analysis. *Communication Monographs*, *46*. doi: 10.1080/03637757909375987

Nunamaker, J. F., Derrick, D. C., Elkins, A. C., Burgoon, J. K., & Patton, M. W. (2011, July). Embodied conversational agent-based kiosk for automated interviewing. *Journal of Management Information Systems*, *28*(1), 17–48. doi: 10.2753/MIS0742 -1222280102

Parle, M., Maguire, P., & Heaven, C. (1997). The Development of a training model to improve health professionals' skills, self-efficacy and outcome expectancies when communicating with cancer patients. *Social Science & Medicine*, *44*(2), 231–240.

Parsons, T. D., & Rizzo, A. a. (2008, September). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A Meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, *39*(3), 250–61. doi: 10.1016/j.jbtep.2007.07.007

Pataki, C., Pato, M. P., Sugar, J., Rizzo, A., Parsons, T. D., George, C., & Kenny, P. (2012). Virtual patients as novel teaching tools in psychiatry. *Academic Psychiatry*, *36*(5), 64. doi: 10.1176/ appi.ap.10080118

Pelachaud, C. (2005). Multimodal expressive embodied conversational agents. In *MM'05* (pp. 683–689). ACM. Retrieved from `http://citeseerx.ist.psu.edu/viewdoc/ download?doi=10.1.1.135.6494{&}rep=rep1{&}type=pdf`

Pelachaud, C. (2009). Modelling multimodal expression of emotion in a virtual agent. *Philosophical Transactions of the Royal Society of London*, *364*, 3539–3548. doi: 10.1098/rstb.2009.0186

Perotti, V. S., & DeWine, S. (1987). Competence in communication: An Examination of three instruments. *Management Communication Quarterly*, *1*(2), 272–287.

Pertaub, D.-P., Slater, M., & Barker, C. (2001, jan). An Experiment on fear of public speaking in virtual reality. *Studies in Health Technology and Informatics*, *81*, 372–8. Retrieved from `http:// www.ncbi.nlm.nih.gov/pubmed/11317771`

Pertaub, D.-P., Slater, M., & Barker, C. (2002). An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators & Virtual Environments*, *11*(1), 68–78. doi: 10.1162/105474602317343668

Plutsky, S. (1996). Faculty perceptions of students' business communication needs. *Business Communication Quarterly*, *59*(4), 69–76. doi: 10.1177/108056999605900407

Poeschl, S., & Doering, N. (2012, mar). Virtual training for fear of public speaking: Design of an audience for immersive virtual environments. *2012 IEEE Virtual Reality (VR)*, 101–102. Retrieved from `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htmlarnumber=6180902` doi: 10.1109/VR.2012.6180902

Poeschl, S., Wall, K., & Doering, N. (2013). Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence. In *Virtual Reality (VR), 2013 IEEE* (pp. 129–130).

Poyatos, F. (1985). The Deeper levels of face-to-face interaction. *Language and Communication*, *5*(2), 111–131. doi: 10.1016/0271-5309(85)90004-7

Price, M., & Anderson, P. (2007). The role of presence in virtual reality exposure therapy. *Journal of Anxiety Disorders*, *21*(5), 742–751. doi: 10.1016/j.janxdis.2006.11.002

Putnam, L. L. (1979). Role functions and role conflicts of communication trainers. *Journal of Business Communication*, *17*(1), 37–52. doi: 10.1177/002194367901700104

Raij, A. B., Johnsen, K., Dickerson, R. F., Lok, B. C., Cohen, M. S., Duerson, M., ... Lind, D. S. (2007). Comparing interpersonal interactions with a virtual human to those with a real human. *IEEE Transactions on Visualization and Computer Graphics*, *13*(3), 443–457. doi: 10.1109/TVCG.2007.1036

Rank, S., & Petta, P. (2012). Backstory authoring for affective agents. In *5th International Conference, ICIDS 2012* (pp. 144–149). San Sebastian. doi: 10.1007/978-3-642-34851-8{_}14

Rapee, R., & Heimberg, R. G. (1997). A Cognitive-behavioral model of anxiety in social phobia. *Behavior Research and Therapy*, *35*(8), 741–756.

Ready, D. J., Pollack, S., Rothbaum, B. O., & Alarcon, R. D. (2006). Virtual reality exposure for veterans with posttraumatic stress disorder. *Journal of Aggression, Maltreatment & Trauma*, *12*(1-2), 199–220. Retrieved from `http://www.tandfonline.com/doi/abs/10.1300/J146v12n01{_}07` doi: 10.1300/J146v12n01

Redish, J. (2007). Expanding usability testing to evaluate complex systems. *Journal of Usability Studies*, *2*(3), 102–

111. Retrieved from `https://uxpa.org/sites/default/files/utesting-complex{_}0.pdf`

Remington, R. W., Johnston, J. C., & Yantis, S. (1992). Involuntary attentional capture by abrupt onsets. *Perception & Psychophysics*, *51*(3), 279–290. doi: 10.3758/BF03212254

Rettie, R. (2004). Using Goffman's frameworks to explain presence and reality. In *7th Annual International Workshop on Presence* (pp. 117–124). Valencia, Spain. Retrieved from `http://www.temple.edu/ispr/prev{_}conferences/proceedings/2004/index.html{#}ConfPgm`

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. Retrieved from `http://www.as.wvu.edu/{~}richmond/articles/ch6-richmond-teacher.pdf`

Rist, T., André, E., Baldes, S., Gebhard, P., Klesen, M., Kipp, M., ... Schmitt, M. (2003). A Review on the development of embodied presentation agents and their application fields. In H. Prendinger & M. Ishizuka (Eds.), *Life-like character: Tools, affective functions, and applications* (pp. 377–404). Berlin Heidelberg: Springer-Verlag. Retrieved from `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.83.9484{&}rep=rep1{&}type=pdf`

Riva, G. (2000). Design of clinically oriented virtual environments: A Communicational approach. *CyberPsychology & Behavior*, *3*(3), 351–357. Retrieved from `http://proxy.govst.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true{&}db=psyh{&}AN=2000-16537-006{&}site=ehost-live` doi: 10.1089/10949310050078797

Riva, G., Gaggioli, A., Grassi, A., Raspelli, S., Cipresso, P., Pallavicini, F., ... Donvito, G. (2011). NeuroVR 2–A free virtual reality platform for the assessment and treatment in behavioral health care. *Studies in Health Technology and Informatics*, *163*, 493–495. doi: 10.3233/978-1-60750-706-2-493

Rizzo, A. (2008). Virtual reality in psychology and rehabilitation: The Last ten years and the next! In *Proceeding of the 7th ICDVRAT with ArtAbilitation* (pp. 3–6). Maia, Portugal.

Rizzo, A., Bowerly, T., Buckwalter, J. G., Schultheis, M., Matheis, R., Shahabi, C., ... Sharifzadeh, M. (2002). Virtual environments for the assessment of attention and memory processes: The Virtual

classroom and office. , 3–12.

Rizzo, A., Kenny, P., & Parsons, T. D. (2011). Intelligent virtual patients for training clinical skills. *Journal of Virtual Reality and Broadcasting*, *8*(3), 1–16. Retrieved from `https://www.jvrb.org/past-issues/8.2011/2902`

Robillard, G., Bouchard, S., Fourier, T., & Renaud, P. (2003). Anxiety and presence during VR immersion: A Comparative study of the reactions of phobic and non-phobic participants in therapeutic virtual environments derived from computer games. *Cyberpsychology & Behavior*, *6*(5), 467–481.

Rogers, E. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.

Rogers, P. (1988, aug). Distinguishing public and presentational speaking. *Management Communication Quarterly*, *2*(1), 102–115. Retrieved from `http://mcq.sagepub.com/cgi/doi/10.1177/0893318988002001007` doi: 10.1177/0893318988002001007

Rothbaum, B. O., Anderson, P., Zimand, E., Hodges, L., Lang, D., & Wilson, J. (2006). Virtual reality exposure therapy and standard (in vivo) exposure therapy in the treatment of fear of flying. *Behavior Therapy*, *37*(1), 80–90. doi: 10.1016/j.beth.2005.04.004

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. doi: 10.1037/0022-006x.68.6.1020

Ruddle, R. A., Payne, S. J., & Jones, D. M. (1999). Navigating large-scale virtual environments: What differences occur between helmet-mounted and desk-top displays? *Presence: Teleoperators and Virtual Environments*, *8*(2), 157–168.

Safir, M. P., Wallach, H. S., & Bar-Zvi, M. (2012, March). Virtual reality cognitive-behavior therapy for public speaking anxiety: One-year follow-up. *Behavior Modification*, *36*(2), 235–46. doi: 10.1177/0145445511429999

Saldaña, J. (2010). *The Coding manual for qualitative researchers.* Sage.

Saleh, N. (2010). The Value of virtual patients in medical education. *Annals of Behavioral Science and Medical Education*, *16*(2), 29–31.

Salmon, P., & Young, B. (2011, mar). Creativity in clinical communication: From communication skills to skilled communication.

Medical Education, *45*(3), 217–26. doi: 10.1111/j.1365-2923.2010 .03801.x

Sánchez, Á., Barreiro, J., & Maojo, V. (2000). Design of virtual reality systems for education: A Cognitive approach. *Education and Information Technologies*, *5*(4), 345–362. Retrieved from `http:// link.springer.com/article/10.1023/A:1012061809603`

Sanchez-Vives, M. V., & Slater, M. (2005, April). From presence to consciousness through virtual reality. *Nature reviews. Neuroscience*, *6*(4), 332–9. Retrieved from `http://www.ncbi.nlm.nih.gov/ pubmed/15803164` doi: 10.1038/nrn1651

Scherer, K. R. (1994). Affect bursts. In S. van Goozen, N. Van de Poll, & J. Sergeant (Eds.), *Essays on emotion theory* (1st ed., pp. 161–196). New Jersey: Lawrence Erlbaum.

Scherer, K. R., & Ellgring, H. (2007, February). Multimodal expression of emotion: Affect programs or componential appraisal patterns? *Emotion (Washington)*, *7*(1), 158–71. doi: 10.1037/1528-3542.7 .1.158

Schreier, M. (2012). *Qualitative content analysis in practice*. Los Angeles: Sage.

Schutt, R. K. (2014). *Fundamentals of social work research* (Second ed.; R. G. Engel & R. K. Schutt, Eds.). Los Angeles: Sage.

Seron, F. J., Gutierrez, D., Magallon, J. A., Sobreviela, E. J., & Gutierrez, J. A. (2004, apr). A CAVE-like environment as a tool for full-size train design. *Virtual Reality*, *7*(2), 82–93. doi: 10.1007/s10055-003-0117-6

Sharma, S., Otunba, S., & Han, J. (2011). Crowd simulation in emergency aircraft evacuation using virtual reality. In *The 16th international conference on computer games* (pp. 12–17).

Sherman, W. R., & Craig, A. B. (2003). *Understanding virtual reality: Interface, application, and design.* San Francisco: Morgan Kaufmann.

Shibata, T. (2002). Head mounted display. *Displays*, *23*(1-2), 57–64. doi: 10.1016/S0141-9382(02)00010-0

Silverman, N. I., Barry Gand Badler, Pelechano, N., & O'Brien, K. (2005). *Crowd simulation incorporating agent psychological models, roles and communication.*

Slater, M. (2003). *A Note on presence terminology* (Vol. 3).

Slater, M. (2009, dec). 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), 3549–57. doi: 10.1098/rstb.2009.0138

Slater, M., Linakis, V., Usoh, M., & Kooper, R. (1996). Immersion, presence, and performance in virtual environments: An Experiment with tri-dimensional chess. In *ACM Virtual Reality Software and Technology* (pp. 163–172). Retrieved from `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.34.6594{&}rep=rep1{&}type=pdf`

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. doi: 10.1109/38.749116

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. Retrieved from `http://discovery.ucl.ac.uk/79956/` doi: 10.1007/s10750-008-9541-7

Smith, D. (2003). Five principles of research ethics. *Monitor on Psychology*, *34*(1).

Smith, E. R., Miller, D. A., Maitner, A. T., Crump, S. A., Garcia-Marques, T., & Mackie, D. M. (2006). Familiarity can increase stereotyping. *Journal of Experimental Social Psychology*, *42*(4), 471-478. doi: http://dx.doi.org/10.1016/j.jesp.2005.07.002

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, *42*(4), 73–93. Retrieved from `http://doi.wiley.com/10.1111/j.1460-2466.1992.tb00812.x` doi: 10.1111/j.1460-2466.1992.tb00812.x

Stevens, A., Hernandez, J., Johnsen, K., Dickerson, R., Raijb, A., Harrison, C., . . . Lind, D. S. (2006). The Use of virtual patients to teach medical students history taking and communication skills. *The American Journal of Surgery*, *191*(6), 806–811.

Street, R. L., & De Haes, H. C. J. M. (2013, October). Designing a curriculum for communication skills training from a theory and evidence-based perspective. *Patient education and counseling*, *93*(1), 27–33. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/23890580` doi: 10.1016/j.pec.2013.06.012

Stuevel, S. A., Ennis, C., & Egges, A. (2014). Mass population: Plausible and practical crowd simulation. In M. C. Angelides & H. Agius (Eds.), *Handbook of digital games* (First ed., pp. 146–174). John Wiley & Sons. doi: 10.1002/9781118796443.ch6

Sutherland, I. E. (1968). A Head-mounted three dimensional display. *Proceedings Fall Joint Computer Conference*, 757–764. Retrieved

from `http://dl.acm.org/citation.cfm?id=1476589.1476686` doi: 10.1145/1476589.1476686

Swartout, W., Traum, D., Artstein, R., Noren, D., Debevec, P., Bronnenkant, K., ... White, K. (2010). Ada and Grace: Toward realistic and engaging virtual museum guides. In *Intelligent virtual agents: Lecture notes in computer science* (Vol. 6356, pp. 286–300). Springer Berlin Heidelberg. doi: 10.1007/978-3-642-15892-6{_}30

Tortella-Feliu, M., Botella, C., Llabrés, J., Bretón-López, J. M., del Amo, R. A., Baños, R. M., & Gelabert, J. M. (2011). Virtual reality versus computer-aided exposure treatments for fear of flying. *Behavior Modification*. doi: 10.1177/0145445510390801

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)*.

Turk, C. (2004). *Effective speaking: Communicating in speech* (2nd ed.). Spon Press. Retrieved from `http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract`

Ulicny, B., & Thalmann, D. (2001). Crowd simulation for interactive virtual environments and VR training systems. In *Proceedings of the eurographic workshop on computer animation and simulation*.

Ulicny, B., & Thalmann, D. (2002). Towards interactive real-time crowd behavior simulation. , *21*(4), 767–775.

Västfjäll, D. (2003). The Subjective sense of presence, emotion recognition, virtual environments. *Cyberpsychology & Behavior*, *6*(2), 181–189.

Verderber, R. F., Sellnow, D. D., & Verderber, K. S. (2010). *The Challenge of effective speaking* (15th ed.). Boston: Wadsworth Cengage Learning.

Verderber, R. F., Verderber, K. S., & Sellnow, D. D. (2008). *The Challenge of effective speaking.* Belmont: Thomson Wadsworth.

Vickers, A. (1997). How to ... give a talk at a conference. *Complementary Therapies in Medicine*, *5*(3), 175–177. doi: 10.1016/S0965-2299(97)80063-5

Vogel-Walcutt, J. J., Fiorella, L., Carper, T., & Schatz, S. (2011, oct). The Definition, assessment, and mitigation of state boredom within educational settings: A Comprehensive review. *Educational Psychology Review*, *24*(1), 89–111. doi: 10.1007/s10648-011-9182-7

von Laban, R. (1980). *The Mastery of movement* (L. Ullmann, Ed.).

Macdonald and Evans.

Vrij, A., Semin, G. R., & Bull, R. (1996). Insight into behavior displayed during deception. *Human Communication Research*, *22*(4), 544–562. doi: 10.1111/j.1468-2958.1996.tb00378.x

Wallach, H. S., Safir, M. P., & Bar-Zvi, M. (2009, May). Virtual reality cognitive behavior therapy for public speaking anxiety: a randomized clinical trial. *Behavior Modification*, *33*(3), 314–38. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/19321811` doi: 10.1177/0145445509331926

Wallach, H. S., Safir, M. P., & Bar-Zvi, M. (2011, January). Virtual reality exposure versus cognitive restructuring for treatment of public speaking anxiety: A Pilot study. *The Israel Journal of Psychiatry and Related Sciences*, *48*(2), 91–7. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/22120443`

Wallbott, H. G. (1998, November). Bodily expression of emotion. *European Journal of Social Psychology*, *28*(6), 879–896. doi: 10.1002/(SICI)1099-0992(1998110)28:6<879::AID-EJSP901>3.0.CO;2-W

Waly, A. F., & Thabet, W. Y. (2002). A Virtual construction environment for preconstruction planning. , *12*, 139–154.

Wang, R., & Bowman, D. A. (2007). Quantifying the Benefits of Immersion for Procedural Training.

Watzlawick, P., Beavin, J. H., & Jackson, D. D. (1967). Some tentative axioms of communication. In *Pragmatics of human communicaiton: A study of interactional pattersn, pathologies, and paradoxes* (pp. 48–71). New York: W.W. Norton & Company.

Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology*, *78*(1), 105–119. doi: 10.1037//0022-3514.78.1.105

Westenberg, P. M., Bokhorst, C. L., Miers, A. C., Sumter, S. R., Kallen, V. L., van Pelt, J., & Blöte, A. W. (2009, October). A Prepared speech in front of a pre-recorded audience: Subjective, physiological, and neuroendocrine responses to the Leiden Public Speaking Task. *Biological Psychology*, *82*(2), 116–24. doi: 10.1016/j.biopsycho.2009.06.005

Whitelock, D., Romano, D., Jelfs, A., & Brna, P. (2000). Perfect presence: What does this mean for the design of virtual learning environments? *Education and Information Technologies*, *5*(4),

277–289. Retrieved from `http://link.springer.com/article/10.1023/A:1012001523715$\delimiter"026E30F$nhttp://link.springer.com/article/10.1023/A:1012001523715$\delimiter"026E30F$nhttp://link.springer.com/content/pdf/10.1023/A:1012001523715.pdf` doi: 10.1023/A:1012001523715

Wiederhold, B. K., & Bouchard, S. (2014). *Advances in virtual reality and anxiety disorders*. London: Springer. doi: 10.1016/S1571-5078(08)00422-4

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. Retrieved from `http://fearofflyingexpert.com/wp-content/uploads/2013/07/Treatment-of-the-Fear-of-Flying{_}2002.pdf`

Wiederhold, B. K., & Wiederhold, M. D. (2005). How treatment is conducted at the virtual reality medical center. In *Virtual reality therapy for anxiety disorders: Advances in evaluation and treatment* (chap. 4). Washington, DC: American Psychological Association. Retrieved from `http://csaweb108v.csa.com.proxy.lib.wayne.edu/ids70/view{_}record.php?id=3{&}recnum=339{&}log=from{_}res{&}SID=r6b4cpqvhi77fq1a0g8dsn4ud0`

Wiemann, J. M. (1981). Effects of laboratory videotaping procedures on selected conversation behaviors. *Human Communication Research*, *7*(4), 302–311.

Wilson, K., & Korn, J. H. (2007). Attention during lectures: Beyond 10 minutes. *Teaching of Psychology*, *34*(2), 85–89. doi: 10.1177/009862830703400202

Winn, W. (1993). *A Conceptual basis for educational applications of virtual reality* (Tech. Rep.). Seattle: University of Washington.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A Presence questionnaire. *Presence: Teleoperators*, *7*(3), 225–240. doi: 10.1162/105474698565686