

Affective and Attentive Interaction with Virtual Humans in Gaze-based Settings

DISSERTATION

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften

eingereicht an der

Fakultät für Angewandte Informatik
Universität Augsburg

von

Dipl.-Inf. Nikolaus Bee

Januar 2013

Gutachter: Prof. Dr. Elisabeth André
Prof. Dr. Marc Erich Latoschik
Prof. Dr. Bernhard Möller

Mündliche Prüfung: 08.01.2013

Danksagung

Mein besonderer Dank gilt meiner Betreuerin Frau Prof. Dr. Elisabeth André, die mir nicht nur diese Promotion ermöglicht hat, sondern auch in herausfordernden Situation immer unterstützend zur Seite stand. Auch gilt mein Dank Herrn Prof. Dr. Marc Erich Latoschik und Herrn Prof. Dr. Bernhard Möller, die diese Dissertation spontan mit ihren hilfreichen Hinweisen und Vorschlägen unterstützt haben.

Dank gilt auch meinen Kolleginnen und Kollegen am Lehrstuhl für Human-Centered Multimedia für die gute Zusammenarbeit und die vielen nützlichen Gespräche. Besonders möchte ich hier Michael Wißner für die Zusammenarbeit in der Spieleprogrammierung und Weiterentwicklung der Horde3D GameEngine danken. Aber natürlich auch Ionut Damian, Felix Kistler und Dominik Sollfrank, die die Entwicklung der Horde3D GameEngine tatkräftig unterstützt haben.

Desweiteren möchte ich mich bei allen Studenten bedanken, die in den Lehrveranstaltungen, Abschlussarbeiten oder als Hilfswissenschaftler meine Arbeiten unterstützt haben. Besonders danke ich hier der Susanne Tober, Bernhard Falk, Nicolas Schulz, Stefan Franke und Kathrin Janowski.

Zudem möchte ich mich bei Prof. Dr. Helmut Prendinger bedanken, der durch meinen Aufenthalt in Japan nicht nur meinen Blick in kultureller Hinsicht erweitert hat. Auch Fred Charles, Diana Arellano und Marilyn Walker möchte ich für ihre Zusammenarbeit danken.

Ein ganz besonderer Dank gilt meiner Frau Karin, die dann letztendlich immer wieder den notwendigen Ehrgeiz in mir geweckt hat.

Abstract

This thesis investigates the affective and attentive gaze-based interaction with virtual humans and will answer related questions to affect and attention and the challenges facing the generation of an anthropomorphic interface for human-machine interaction. One focus is on recommendations for the generation of gaze aware interfaces, the control and perception of virtual agent's facial expressions and an interactive gaze model for the face-to-face interaction with a virtual agent. Our focus is not on the analysis of the semantics, but on the socio-emotional aspects of such a conversation.

This thesis gives a general view on the different functions of gaze and eye contact in human-human interaction and related work of human-agent interaction and how interfaces with explicit and implicit gaze interaction could look like and reveal the challenges of such interfaces. While facial expression control can be quite complicated, this thesis demonstrates how facial expression control can be simplified.

The investigation of the perception of facial displays inbetween humans and virtual humans is part of this thesis. It investigates how humans perceive variations of the facial display of a virtual human. One part deals with how humans perceive the emotional component of the combination of affective facial expressions, head direction and eye orientation. Another part investigates whether the head direction and eye orientation influence the perception of the personality of the virtual human.

Finally, mutual gaze plays an important role to establish attention with an interlocutor. This thesis deals with a gaze model that is aware of a user's current gaze with the help of an eye tracker. It analyzes where a user is looking and if the user is looking into the eyes of a virtual human to establish mutual gaze. It investigates if a human recognizes the interactive gaze model in a non-verbal setting and whether a human recognizes it in a verbal setting.

Zusammenfassung

Diese Arbeit beschäftigt sich mit der affektive Interaktion und Aufmerksamkeit beim Blicken zwischen virtuellen und realen Menschen. Sie beantwortet dazugehörige Fragen zum Thema Affekt und Aufmerksamkeit und zeigt die Herausforderungen auf die man bei der Erstellung von anthropomorphen Schnittstellen in der Mensch-Maschine Interaktion trifft. Ein Schwerpunkt liegt in Empfehlungen für die Erzeugung von blickaktivierten Interaktionsschnittstellen, ein weiterer in der Steuerung und Wahrnehmung von Gesichtsausdrücken eines virtuellen Agenten und schließlich zeigt sie, wie man ein interaktives Blickmodell für die Interaktion zwischen Mensch und virtuellen Agenten erstellt.

Diese Arbeit gibt einen generellen Blick auf die unterschiedlichen Funktionen von Blickkontakt in Mensch-Mensch Interaktion und führt in die verwandten Arbeiten zum Thema Mensch-Agenten Interaktion ein. Es wird an jeweils einem Beispiel gezeigt, wie man blickbasierte Schnittstellen erzeugt, die zum einen aktives Verhalten und zum anderen passives Verhalten von einem Anwender voraussetzen. Darüberhinaus beschäftigt sie sich, wie man die Steuerung von Gesichtsausdrücken virtueller Agenten vereinfachen kann.

Teil dieser Arbeit ist auch die Wahrnehmung von Gesichtsausdrücken zwischen Menschen und virtuellen Agenten zu untersuchen. Es wird untersucht, wie Menschen Gesichtsausdrucksvariationen von virtuellen Agenten wahrnehmen. Ein Teil geht um die Wahrnehmung von emotionalen Gesichtsausdrücken in Kombination mit verschiedenen Kopforientierungen und Blickrichtungen. Ein weiterer Teil untersucht, ob Kopfrichtungen die Wahrnehmung von Persönlichkeit virtueller Agenten beeinflusst.

Schließlich wird noch ein Blickmodell untersucht, das Blickkontakt zu einem Nutzer aufbauen kann. Mit Hilfe eines Eye Trackers ist das System in der Lage den gegenwärtigen Blick eines Anwenders zu erkennen und den Blick eines virtuellen Agenten entsprechend steuern. Es wird analysiert wo ein Anwender im Augenblick hinsieht und ob er dem virtuellen Agenten direkt in die Augen sieht. Dazu wird untersucht, ob ein Mensch solch ein interaktives Blickmodell wahrnimmt und ob es Unterschiede zwischen einem nicht-verbalen und verbalen Interaktionsszenario gibt.

Contents

1	Introduction	1
1.1	Research Questions	2
1.2	Content Overview	4
2	Background	5
2.1	Gaze Interaction in Graphical User Interfaces	5
2.1.1	Gaze-Controlled Selection	7
2.1.2	Gaze-Controlled Gesturing	8
2.1.3	Gaze-Controlled Continuous Selection	9
2.2	Gaze in Human-Human Interaction	10
2.3	Gaze in Human-Agent Interaction	12
2.3.1	Humanoid Virtual Agents	12
2.3.2	Gaze during Non-Verbal Interaction	15
2.3.3	Gaze during Verbal Interaction	17
3	Study I: Explicit vs. Implicit Gaze Interaction	21
3.1	Experiment I: Explicit Interaction	21
3.1.1	Implementation of Eye Writing Applications	22
3.1.2	Experiment	25
3.1.3	Results	25
3.2	Experiment II: Implicit Interaction	28
3.2.1	Gaze Cascade Effect	28

3.2.2	Study	29
3.2.3	Results	31
3.3	Conclusion	31
4	Framework for Gaze-Based Human-Agent Interaction	33
4.1	Horde3D GameEngine	33
4.1.1	Facial Expressions	35
4.1.2	Lip Synchronization	37
4.1.3	Inverse Kinematics for Gaze	38
4.1.4	3D Object Detection	39
4.2	Experiment: Facial Expression Control for Virtual Characters	41
4.2.1	Control of Facial Expressions	41
4.2.2	FACS-based Facial Expression Generation	43
4.2.3	Design and Implementation of the Interface	44
4.2.4	Studies with Professional Graphics Designers	52
4.2.5	Evaluation of the Hardware Controllers	53
4.3	Conclusion	58
5	Study II: Perception of Facial Display, Head and Eye Orientation	61
5.1	Experiment I: Dominance Perception	62
5.1.1	Affective and Attentive Virtual Character	63
5.1.2	Experimental Study	64
5.1.3	Results	65
5.2	Experiment II: Personality Perception – Extraversion, Agreeable- ness and Emotional Stability	73
5.2.1	Experimental Study	74
5.2.2	Results	78
5.3	Conclusion	83

6	Study III: Gaze Interaction with Virtual Characters	87
6.1	Eye Orientation in Human-Agent Interaction	88
6.2	Interactive Gaze Model for Human-Agent Interaction	89
6.3	Experiment I: Non-Verbal Interaction	91
6.3.1	Gaze Model for Human-Agent Interaction	92
6.3.2	System	96
6.3.3	Evaluation	97
6.4	Experiment II: Verbal Interaction	103
6.4.1	Analysis of Conversational and Social Behaviors	104
6.4.2	Virtual Character	107
6.4.3	Evaluation of the Gaze Models	107
6.5	Conclusion	113
7	Conclusion	115
7.1	Summary	115
7.2	Contributions	118
7.2.1	Methodological Contributions	118
7.2.2	Theoretical Contributions	119
7.2.3	Practical Contributions	119
7.3	Future Work	120
A	Facial Expression XML	121
B	Horde3D GameEngine Configuration for Alfred	125
C	Questionnaires	129
C.1	Facial Expression Control	129
C.2	Perception of Dominance	131
C.3	Perception of Personality	132
C.4	Non-Verbal Interaction	134
C.5	Verbal Interaction	137
	Bibliography	141

List of Figures

3.1	Adapted interface of Quikwriting for use with gaze. The dotted line indicates the gaze path to write the letter 'g'. The user currently looks at 'g'. The light shaded background follows the gaze and indicates which section the user is looking at.	24
3.2	In a new design of the adapted Quikwriting, we will place the text field in the resting area. This will enable users to check what they have written without moving their eyes to any place outside the control interface.	27
3.3	The user sits in front of a 19 screen and an eye tracker. The system automatically selects the preferred tie dependent on the user's gaze behavior.	30
4.1	The game application (top) interacts with the game world (bottom) containing several entities (e.g. Marie or Crate1) each having one or more components (e.g. Physics or TTS).	34
4.2	Visualization of the picking objects for detecting the eyes and the head of Alfred.	40
4.3	The virtual character Alfred is designed utilizing FACS to compose facial expressions.	44
4.4	The XBox 360 controller with two analog sticks, two analog shoulder buttons, one digital stick and several buttons.	45
4.5	The P5 data glove with five analog controllers and three buttons.	47
4.6	Settings for the gamepad to control the action units for the upper face (left), lower face without inner lips (middle), and the inner lips (right).	48
4.7	Facial expression without (left) and with (right) constraint model.	51

4.8	Subjective user ratings for the Data Glove compared to the Slider based approach (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).	55
4.9	Subjective user ratings for the Gamepad compared to the Slider based approach (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).	57
5.1	The six facial expressions of Alfred used in the study. (Upper row from left to right: joy, anger and sadness; lower row: fear, disgust and neutral)	63
5.2	Dominance values with all eye and head orientation variations . . .	65
5.3	Dominance values with eyes and head directed at the user	67
5.4	Dominance values for anger with directed head orientation and different gaze directions	69
5.5	Dominance values for anger and head orientation	70
5.6	Dominance values for neutral and head orientation	71
5.7	The virtual character Alfred.	74
5.8	Mean values for <i>extraversion</i> dependent on the head orientation. . .	79
5.9	The head orientation (<i>center down</i>) with the lowest rating (left) and the one (<i>up side</i>) with the highest rating (right) for <i>Extraversion</i>	80
5.10	Mean values for <i>agreeableness</i> dependent on the head orientation. .	81
5.11	The head orientation (<i>center up</i>) with the lowest rating (left) and the one (<i>center down</i>) with the highest rating (right) for <i>Agreeableness</i> . . .	82
5.12	Mean values for <i>Emotional Stability</i> dependent on the head orientation.	83
5.13	The head orientation (<i>center up</i>) with the lowest rating (left) and the one (<i>center side</i>) with the highest rating (right) for <i>emotional stability</i> . .	84
6.1	Variations of the eye angle: (1) focusing a point in front of the subject, (3) focusing the subject, (5) parallel eye direction.	89
6.2	The charts show the results for the focus experiment (left: looking like a real person, right: seeing through me).	90
6.3	The gaze model reacts on mutual gaze [MG] between user and virtual character. If a mutual gaze is recognized, the state switches from Gaze at user to Gaze averted for a specific time.	91

6.4	The <i>attention phase</i> models the phase in which men and women arise each other's attention. It is characterized by ambivalent non-verbal behavior, such as a brief period of mutual gaze broken by downward eye aversion, reflecting the uncertainty of the first seconds.	94
6.5	In the <i>recognition phase</i> , one person recognizes the interest of the other. He or she may then discourage the other person, for example, by a downward gaze, or signal readiness to continue the interaction, for example, by a friendly smile.	95
6.6	The virtual character Alfred with a skyline in the background. . . .	96
6.7	Set-up for the gaze based interaction application from different perspectives.	98
6.8	Results of the questionnaire for the interactive, non-interactive ideal and non-interactive anti-flirt mode ($*p < 0.05$, $**p < 0.01$).	101
6.9	We measure user interaction with different sensor devices, which are synchronized and pre-processed through the framework.	105
6.10	Set-up for the interaction with Emma.	108
6.11	Results for the questions compared with the <i>interactive</i> and <i>non-interactive</i> gaze model while interacting with Emma ($*p < 0.05$). . .	110
6.12	Gaze pattern before the users start speaking. The vertical axis indicates the gaze target(0 = looking away, 1 = looking at Emma, <i>red line</i> = average) during conversation. The user starts speaking at $t = 0$.112	
6.13	Gaze pattern after the users stopped speaking. The vertical axis indicates the gaze target(0 = looking away, 1 = looking at Emma). The user stops speaking at $t = 0$	113

List of Tables

3.1	Comparison of writing speeds in wpm (words per minute)	26
5.1	Comparison between Russell and Mehrabian's and our dominance values	66
5.2	Post-hoc comparisons for averted gaze and head orientation between different emotions (** $p = 0$, n.s. = not significant)	68
5.3	Median values for dominance over all emotions with directed eye and head direction compared with averted eye and head direction .	68
5.4	Results of the perception test for emotional expressions of Alfred. Multiple nominations were allowed.	72
5.5	Varying head orientation.	76
5.6	Questionnaire items for perception of head orientation.	77
5.7	Post-hoc comparisons for <i>Extraversion</i> and the varying head orientations (U = up, C = center, D = down, S = side, M = middle, $^+p < 0.1$, $^*p < 0.05$, $^{***}p < 0.001$, n.s. = not significant).	80
5.8	Post-hoc comparisons for <i>agreeableness</i> and varying head orientation (U = up, C = center, D = down, S = side, M = middle, $^{**}p < 0.01$, $^{***}p < 0.001$, n.s. = not significant)	82

Chapter 1

Introduction

Der Blick in das Auge des Andern dient nicht nur mir, um jenen zu erkennen, sondern auch ihm, um mich zu erkennen; auf der Linie, die beide Augen verbindet, trägt er die eigne Persönlichkeit, die eigne Stimmung, den eigenen Impuls zu dem Andern hin.

Georg Simmel

Vending machines, like ticket machines, automatic teller machines or self-service check-ins at airports have become surprisingly fast common during the last decades. Except for the security check, it is meanwhile possible to board an aircraft without any personal contact. This takes the ability to being able to handle all these machines for granted. But not only the humans need to be able to handle these machines, the machines also need to be able to understand what the user wants. What happens for example, if the passport at the self-service check-in machine at the airport cannot be read by the machine, because the user is not understanding the instructions of the machine? Or, for example, is it really faster to buy a train ticket at a ticket machine instead of using the old-style ticket counters with personal contact?

While these kind of services shift from human interaction to machine interaction, the interaction with these machines becomes unnatural and functional. The error rate of such machine interactions will be reduced with the development of better systems over time. Still missing, however, is the human-like interaction with such systems. As the technology for natural interfaces like speech recognition, face detection or body tracking improves, it might be possible to substitute the

menu based interfaces in these machines by virtual humans that act and react like real humans. Such anthropomorphic interfaces could offer all the advantages of a human-like interaction.

While in the past we had personal interaction with real humans, we currently are using explicit menu-based interaction with machines to buy, for examples, train tickets. Such interfaces could be endowed with gaze control. That would mean that buying a train ticket could be done by explicitly looking at ticket options presented on the display and the user would not have to select these options by touching the display. A next step for such interfaces would be to use implicit gaze interaction. This means that such an interface could recommend options for purchasing a train ticket by simply analyzing the user's gaze. For example, if the user is looking for a longer time at specific conditions, the user might be interested in these conditions and the ticket machine may offer them to the user. Such an implicit interface would be able to detect the needs of the current user by, for instance, analyzing the user's current gaze behavior. In any case it is important for such anthropomorphic interfaces to be authentic and this includes attention towards the users.

1.1 Research Questions

This thesis investigates how to build gaze-based interfaces with the goal to study their potential for gaze-based human-machine interaction. We start by discussing differences between deliberately or explicitly controlled and unconsciously or implicitly controlled gaze-based graphical user interfaces before moving to the main focus of the thesis: gaze-based interaction in anthropomorphic interfaces. For an anthropomorphic gaze-based interface like a virtual human, the latter – implicit gaze interaction – will be more relevant. This, for example, means that among other things such an interface needs to act and react naturally to gaze behavior. Thus, we will also focus on what role plays affect in gaze-based interaction, for example which emotions can be transmitted through gaze and how affect and attention can be detected and generated through gaze. This work will offer a solution on how to generate gaze aware interfaces, how to control a virtual agent's facial expressions for expressing affect and attention and what an interactive gaze model for the face-to-face interaction with a virtual agent should look like. We focus not on the analysis of the semantics, but on the socio-emotional aspects of such interaction.

Gaze Aware Interfaces

- *Explicit Gaze Interaction*: The major issue with direct gaze controlled interfaces lies in the nature of the eyes. The main role of the eyes is to gather information. It might be hard for the user to use the eyes as control device and information source concurrently.
- *Implicit Gaze Interaction*: What do specific gaze patterns look like? Are there specific patterns for specific gaze-based interfaces? How can such patterns be recognized and what are relevant parameters?

Perception of Facial Display

- *Emotion Perception*: How do humans perceive facial expressions and how is this related to gaze direction? How do the representation of emotions and head and eye direction interact with each other? A human that is staring at somebody might be considered as more dominant than a human that is looking downwards. Is this still valid for different emotional facial expressions? Is a virtual human staring at somebody as dominant while being angry than while being happy? Is a happy virtual human perceived the same when looking at somebody or looking away?
- *Personality Perception*: How do different head and eye gaze directions influence the perception of a virtual human's personality? Is, for example, a direct gazing virtual human considered as more extrovert and a virtual human looking down considered as more introvert?

Interactive Gaze Model

In human-human interaction mutual gaze plays an important role, for example, in recognizing the attention of an interlocutor or in regulating the conversational flow. How could an interactive gaze model look like that takes the user's gaze into account? How does mutual gaze influence the interaction with a virtual human? The biggest challenge is to build a real-time model for gaze interaction that is on the one side able to detect the user's current gaze intention and on the other side to generate the appropriate gaze answer.

- *Non-verbal Gaze Interaction*: How does an interactive gaze model for non-verbal interaction influence the interaction with a human?

- *Verbal Gaze Interaction*: How can an interactive gaze model that takes verbal interaction into account be built? Do users still perceive the same interactivity of the gaze behavior compared to a non-verbal model or is the verbal communication with a virtual human overlaying the gaze interaction?

1.2 Content Overview

Chapter 2 gives a general view of the different functions of gaze and eye contact in human-human interaction and related work on human-agent interaction. Further, this chapter shows what interfaces with explicit and implicit gaze interaction could look like and demonstrate the challenges of such interfaces.

Parts of the development of the virtual environment for simulating virtual characters was part of this thesis. **Chapter 3** explains the most important components, i.e. facial expression control, lip synchronization, gaze control and object detection, of the Horde3D GameEngine, that were used to control a virtual human. The second part of this chapter deals with the usage of these components and the evaluation how facial expression control can be simplified.

Chapter 4 deals with the perception of facial displays between humans and virtual humans. This chapter investigates how humans perceive variations of the facial display of a virtual human. The first part investigates how humans perceive the emotional component of the combination of affective facial expressions, head direction and eye orientation. The second part investigates whether the head direction and eye orientation influence the perception of the personality of the virtual human.

Mutual gaze plays an important role to establish attention with an interlocutor. **Chapter 5** deals with a gaze model that is aware of a user's current gaze. It analyzes what a user is looking at and whether the user is looking into the eyes of a virtual human to establish mutual gaze. The first part of Chapter 5 investigates if a human recognizes the interactive gaze model in a non-verbal setting and the second part investigates whether a human recognizes it in a verbal setting.

Chapter 6 concludes this thesis with summarizing the scientific and technical contributions and gives an outlook about future works.

Chapter 2

Background

A profound analysis and understanding of the functions of gaze interaction provides the basis for this research on gaze in graphical human-computer interfaces. The first part of this chapter gives an overview of gaze-based graphical user interfaces (GUIs). The second section is about the role of gaze in human-human interaction. The subsequent section gives an overview of related research that is most relevant to this thesis regarding gaze in human-agent interaction. That section focuses on gaze in verbal and non-verbal human-human interaction.

2.1 Gaze Interaction in Graphical User Interfaces

Eye gaze based interaction can be divided into two categories: active and passive interaction. Duchowski (2002) classifies eye tracking applications into two interactive categories named *selective* and *gaze contingent*. Selective interaction systems use the eye as direct selection and input method. For example, it is quite simple to map the user's gaze to a mouse pointer of a display and the user is fully aware of the mouse pointer control by moving the eyes. In a gaze contingent interaction, however, the user should not be able to directly control the interaction. A gaze contingent system is for example using the user's typical gaze behavior during an interaction. An example would be an advertising display that is able to adjust its advertisement depending on the user's interest (i.e. areas where a user is looking most). Schmidt (2000) defines this interaction method in a broader point of view and in the following sections we present examples for explicit and implicit gaze based interaction systems.

One domain for explicit gaze-based interaction, mainly of interest to physically handicapped people, is gaze-based writing. One challenge in such a writing application is that a human's eye continuously gazes, wanders and normally stops

only for a fraction of a second. Common gaze-based writing systems force the users to dwell upon a specific position for a certain time to trigger a command.

Gaze as an input method for explicit interaction is quite challenging compared to using hands and fingers. The eye is normally used to gain information only and not to trigger any commands or control devices. Ashmore et al. (2005) summarizes four problems which interface developers should consider when using gaze for human-computer interaction. (1) The accuracy of eye tracking devices is limited to about $0.5 - 1^\circ$ visual angle. 1° corresponds approximately to the size of a thumb-nail at arm length (Duchowski, 2007). This restricts the interaction elements in an interface to a certain size. (2) Gaze is recognized with a delay dependent on the frame rate. A 50 Hz system, for instance, incurs delays of 20 ms. When using webcams with 25 Hz, the delay would be 40 ms. (3) Gaze is never perfectly still even if one concentrates on a point. It slightly jitters with flicks less than 1° , small drifts of about $0.1^\circ/\text{s}$ and tremors (tiny, high frequency eye vibrations). (4) The Midas touch problem (Jacob, 1991) leads to ambiguities in gaze-controlled interaction. Eyes are used for seeing and gathering information. Naturally, they follow any salient or moving point of interest. Interfaces that use gaze should thus be carefully designed and not use too many or intrusive elements that could attract the attention of the user's gaze.

Three types of writing can be distinguished: typing, gesturing and continuous writing.

1. *Typing* – can be done on keyboards or e.g. a game controller where the letters on a keyboard interface are selected. All have in common that users have to press a button to print a letter.
2. *Gesturing* – comparable to block lettering. Such input methods may be found in handhelds with pen-based input devices, such as the Graffiti system from Palm. Each input of a letter is separated by a short interruption (lifting of the pen).
3. *Continuous writing* – reduces the interruptions between letters to a minimum. Cursive handwriting comes close to it.

Taking the humans' gaze behavior into consideration, *continuous writing* matches best the requirements of interfaces that utilize gaze for input control. Human gaze is always moving and always 'on', which can be seen as a pen that always draws. We cannot simply switch off our gaze. Whereas for handwriting it is necessary to lift the pen to separate single words, for a gaze driven application, as we cannot

switch off gaze, it would be fascinating to design an interface where switching is not necessary at all. Before presenting our own system, we will discuss applications that were developed for gaze-controlled writing that fall into the categories introduced above.

2.1.1 Gaze-Controlled Selection

There are two kinds of common keyboard writing. In the case of direct writing, a keyboard with the keys arranged in alphabetical order or using a keyboard layout is displayed. Unlike this, multi-tap writing is based on a hierarchical arrangement of letters. Here the letters are grouped on keys together and users have to repeatedly press buttons to get a single letter. This method is frequently used in mobile phones. To adapt common keyboard writing for gaze writing, users have to directly select the keys with their gaze, in a similar way as they would type on a keyboard. They simply type with their eye. Both gaze-controlled direct writing and gaze-controlled multi-tap use dwell time, i.e. users have to fixate a specific point for a specific time, to trigger a key.

Majaranta et al. (2006) used gaze-controlled interfaces for writing. To write letters, users simply must look at the on-screen button for a specific time. As users write with their eyes in a contact-free manner, there is at first no haptic or acoustic feedback which they might be familiar with from typewriters or keyboards. Majaranta et al. (2006) investigated several kinds of typing feedback for users: visual, speech and click noise. The visual feedback was implemented with a small highlighted border around the key which is displayed when the user looks at the key. Furthermore, the size of the character on the key shrinks linearly with the dwell time. On selection, the background of the key changes its color and the key goes down. When speech is used, the letter is simply spoken out as feedback after its selection. The click noise as feedback is self-explanatory. The authors found in their comparison of speech only, click + visual, speech + visual, and visual only, that click + visual enabled the users to write fastest with their eyes. The maximum writing speed they achieved was about 7.5 wpm (words per minute).

Hansen et al. (2001) developed a writing system with a hierarchical structure called GazeTalk. They reduced the approximately 30 on-screen keys, common for gaze-controlled keyboard-based systems, to ten in their gaze-based multi-tap system. They applied two different methods. The version for novice users arranges the letters alphabetically. First, letters and special characters are grouped on four buttons. After selecting a button the single characters are shown on single buttons and can be selected for writing. Whereas a gaze-controlled system can

select a letter with a single step, this system needs two. The version for advanced users automatically predicts the next letter while the user is writing. In prediction mode only the six most likely letters are shown. If the desired letter is not among them, the user must trigger another button to get back to the alphabetical display.

Among systems without any probabilistic letter prediction or word completion, gaze-controlled keyboard-based systems are the fastest as it takes only one step to enter a letter. But such systems must display all letters at once on the screen. Depending on the accuracy of the eye tracker, the buttons need a certain size. And in the end, the writing interface will need a lot of space. With multi-tap systems, the buttons can be larger and are therefore less vulnerable to inaccuracies of the eye tracking system.

Both the direct writing and the multi-tap approach use dwell time to trigger keys. Dwell time strongly depends on the experience of users and thus has an impact on typing speed and error rate. If the chosen dwell time is too short, users will make more mistakes and if the dwell time is too long, users will strain their eyes. Thus, a reasonable trade-off between typing speed and error rate needs to be found. Spakov and Miniotas (2004) developed an algorithm to adjust dwell time in real-time. They found that a dwell time of 700 ms enables the user to type nearly without any wrongly selected keys. (Hansen et al., 2001) used a dwell time of 750 ms for novice users which they decreased after several hours of usage to 500 ms.

2.1.2 Gaze-Controlled Gesturing

Isokoski (2000) developed a system called MDITIM (Minimal Device Independent Text Input Method) for device-independent text input. Originally it was only tested with a touchpad, a trackball, a computer mouse, a game controller and a keyboard. To adopt it for gaze control, practically no changes were necessary. Only a modifier key, which was previously controlled by pressing a button, was replaced by an area-of-interest, at which users had to look to trigger it. MDITIM encodes letters in commands of directions, i.e. a = NSW, b = SEW, c = ESW, d = SWE, and so forth. The codes consist of three or four directions. If a user, for instance, wishes to write 'c', her eyes have to go to the right, then down and finally to the left. Writing 'cab' results in ESWNSWSEW, which comes close to a continuous writing system. Nevertheless, MDITIM is not a real continuous writing system. For example, combinations, such as 'dc', are encoded by SWEESW, which includes two equal codes in a row. If the system shall be able to recognize such a combination, there must be an interruption in between.

EyeWrite developed by Wobbrock et al. (2008) is a pure gesture-based eye writing

system similar to Graffiti for Palm handhelds. It is the first system that uses letter-like gestures for eye input, in contrast to MDITIM which encodes the alphabet. The interface – the authors chose a window size of 400×400 – is aware of five areas: the four corners and the middle. To provide some guidance for the gaze to write gestures, there are points placed in the corners and in the middle. To write a 't' for example, the gaze must move from the upper left corner to the upper right corner, then to the lower right corner and finally to the middle to indicate that the gesture is terminated. Glancing at the corners suffices to draw the gesture. The system works not completely dwell time free as the gaze must stay for a specified time in the middle for segmentation. The authors specified a dwell time of about 250 ms, which corresponds to half of the typical dwell time that systems use or to twice as much as the average fixation time. The usage of this system is rather similar to MDITIM, but with letter-like gestures, it is easier for users to remember the gestures.

A disadvantage of gesture-based typing systems is that users have to learn the gestures by heart or look them up. That makes the systems difficult to use for occasional users. EyeWrite could be easier to use than MDITIM as EyeWrite uses letter-like gestures, which makes the gestures easier to memorize. The authors of MDITIM do not provide a user study with performance measurements. Wobbrock et al. (2008) conducted a longitudinal study about the performance of their system. Novice users wrote from about 2.5 wpm (words per minute) up to about 5 wpm after 14 sessions.

2.1.3 Gaze-Controlled Continuous Selection

Urbina and Huckauf (2007) introduce three dwell-time free eye writing applications. In Iwrite, keys are arranged alphabetically in a rectangular horseshoe shape. To select a letter users must look at the letter and then look outside the shape. The inner area of the horseshoe displays the currently written text. A similar system called StarWrite arranges the letters on a half-circle in the upper part and a display for the written text in the lower part. Looking at a letter enlarges it and its two neighbors. To select a letter one must then 'drag' it to the lower text field. Again all letters must be displayed at once. Thus, this method is space consuming or vulnerable to inaccuracy of eye tracking.

pEYEWrite (Urbina and Huckauf, 2007) is their third concept of dwell-time free writing. Here letters are arranged hierarchically using pie menus with six sections, where each section groups five letters or special characters. Letters are again arranged alphabetically. The pie is further divided into an inner and an outer part.

The letters are displayed in the inner part of the pie and to trigger a selection of a letter, users must gaze at the corresponding section on the outer part. To write a letter, a user first selects the section that contains the intended letter. After that, a new pie menu pops up that contains one single letter in each of the six sections. After the selection, the pie disappears and the user can continue. This system needs two activations to write a letter.

Maybe the most prominent gaze-controlled text entry system for continuous writing is Dasher (Ward and MacKay, 2002). It does not use any static elements in its design. Letters move from the right to the left and as soon as a letter crosses a border it is selected. The letters move continuously as long as the users looks at the letters. At start the letters are arranged vertically on the very right border of the application. As soon as the user looks at a letter the letter starts to enlarge and move to the left. Dasher uses probabilistic prediction of letters and word completion. Both concepts are seamlessly integrated in the interface. The probability of a letter is directly depicted by its size which facilitates its selection.

2.2 Gaze in Human-Human Interaction

Gaze cannot only be used as input method for graphical user interfaces. It furthermore plays a much more important role in human-human interaction. It is not only our visual channel that enables us to see our environment and our interaction partners (Ellsworth and Ludwig, 1972). For example, one of the most important functions for gaze in human-human interaction is to gather feedback or to regulate a conversation's flow. Further, gaze is also an important modality to regulate the appearance of a conversational partner. In western cultures it is, for instance, often considered as a dominant or impolite behavior if an interlocutor stares at somebody for too long during a conversation (see for example (Kleinke, 1986)). Kendon (1967) proposes to distinguish between monitoring, regulatory and expressive gaze functions. He further considers, for instance, the amount of mutual gazes as regulator for the level of emotionality. Leathers (1991) categorizes functions of eye behaviors into five groups. The *gaze function of attention* can be defined by the length, direction and kind of gaze, which indicate the individual's level of interest in an interlocutor. An example for the *persuasive function* is the credibility that comes with direct eye contact, which has impact on the person's competence and trustworthiness. The *regulatory function* is to show the participants who is going to speak, in listening or in speaking mode. But gaze not only regulates a conversational flow, it further provides an *affective function* that enables a speaker or listener to emphasis affective characteristics during a conversation.

The *power function* is, for example, illustrated by staring. A powerful person that stares at somebody is in general considered as dominant. These five functions are described in more detail within the following paragraphs.

Gaze may be considered one of the most important means to indicate attention in a dialog. Kleinke (1986) shows correlations between attention and gaze. For instance, interviewers are evaluated more attentive when their gaze is relatively high. And further, interviewees give shorter responses when an interviewer is not looking at them.

Related to the effects of attentiveness is the interaction between eye contact, distance and affiliation. Argyle and Dean (1965) found that the physical distance between two conversational partners affects the amount and duration of eye contact. Placing two persons closer together resulted in less and shorter eye contacts and pairs with opposite gender showed here the greatest effect. Another study revealed that the persons stood closer to a conversational partner whose eyes were shut. Later Argyle and Ingham (1972) not only confirmed these results, but also further examined the effects of distance, gaze and mutual gaze. They recorded the total gaze for each subject, the total mutual gaze of each pair, the average glance length of each subject and for each pair the average length of mutual glance. In a first experiment the subjects were placed either two or ten feet apart. In a second experiment they were placed either 3 or 6 feet apart. For the first experiment Argyle and Ingham (1972) found that there was a significant increase of all four gaze measures independent from gender combinations of the pairs. In the second experiment they found that opposite gender pairs showed the greatest effect for distance dependent gaze behavior.

Kendon (1967) examined gaze behavior during dyadic interactions. He found that gaze regulates the speaker's role and the speaker often ends the turn with a prolonged gaze at the listener. The listener then starts to take the turn by looking away before starting to speak, which signals the acceptance of the turn exchange. Kendon (1967) assumes that this prolonged gaze is not only signaling the end of the turn but also is for getting feedback, if the interlocutor will start to speak. But not only does gaze regulate the speaker's role. (Argyle and Cook, 1976) found, for example, that humans look about 75 % at interlocutors while listening and 41 % while speaking.

Argyle et al. (1974) examined how persons perceive gaze patterns of interlocutors. They trained confederates with five gaze patterns. The first gaze pattern is not to look at the interlocutor at all (zero gaze), the second was only looking while talking, the third looking while listening, normal gaze and the last one continuous gaze. One reason to choose these five gaze patterns was because the eye

contact rate increases from zero (0 %) to continuous (100 %). A principal component analysis revealed the two factors, which they named activity/potency and liking/evaluation. Argyle et al. (1974) state that those factors relate to the perception of dominant – submissive and warm – cold behavior. They found normal gaze as the most liked and the zero gaze as the lowest on activity/potency (dominance). And further the continuous gaze was rated highest in activity/potency, which relates to dominance. Sander et al. (2007) added to typical averted and directed gaze different facial expressions. They found that the emotions fear and anger were perceived with significantly different intensity dependent on averted or direct gaze. Fear was rated more intense in combination with averted gaze whereas anger was rated more intense in combination with direct gaze.

Adams and Kleck (2005) observed that direct gaze supports the perception of approach-oriented emotions (such as anger and joy) while averted gaze enhances the perception of avoidance-directed emotions (such as fear and sadness).

2.3 Gaze in Human-Agent Interaction

This section deals with gaze in human-agent interaction. The first part will give a literature overview of two kinds of approaches in the research on gaze behavior for virtual agents. The second part of this section concentrates on gaze during non-verbal and verbal interaction.

2.3.1 Humanoid Virtual Agents

The following section will discuss literature on gaze behavior for humanoid virtual agents. The focus lays on studies of realistic gaze behavior, attentive gaze behavior and approaches on how to model gaze behavior for virtual agents.

Studies on the Perception of Gaze Behavior

Garau et al. (2001) as well as Lee et al. (2002) investigate the effect of gaze models inferred from gaze patterns derived from anthropologic literature (e.g. Hall, 1963; Kendon, 1967, ...). Both research teams observed a superiority of inferred gaze behaviors over randomized gaze behaviors. A follow-up study by Garau et al. (2003) focused on the correlation between visual realism and behavioral realism. They found that this model-based approach improved the quality of communication when a realistic avatar was used compared to, for example, random gaze behavior. For cartoonish avatars, such an effect could not be observed.

Garau et al. (2001) examined the effect of gaze in human-agent interaction and compared the four channels audio-only, random gaze with a virtual agent, gaze with a virtual agent inferred from human-human gaze behavior and video stream of a real person. They measured the perception of the four different channels with four dimensions. The first dimension was named face-to-face and measured to which extent the subjects perceived the interaction as a real face-to-face conversation. The other measurements were involvement, co-presence and partner evaluation, which is the attitude towards the channel and to what extent the conversation was enjoyed. Garau et al. (2001) found that in general the video condition with the interaction of a real human outperformed the other three conditions as expected. Within the two virtual agent conditions, the one with inferred gaze always outperformed the condition with random gaze and except in the measurement of co-presence also the audio-only condition. This leads to the conclusion that a virtual agent with informed gaze behavior is able to improve human-agent interaction. In a follow-up study Garau et al. (2003) and Vinayagamoorthy et al. (2004) examined how gaze behavior on virtual agents is perceived in an immersive virtual environment with low and high visual quality virtual agents. They found that the inferred gaze behavior outperformed the random gaze behavior for the high quality virtual agent, while the random gaze behavior was perceived better for the low quality virtual agent. This indicates that the visual quality of a virtual agent needs to increase with the quality of realistic gaze behavior.

Lee et al. (2002) investigated how gaze behavior for an virtual agent can be generated from real eye data recorded with an eye tracker. Their eye movement model combined of a model for talking mode and listening mode is based on empirical data for the saccades and fixations of the eye behavior and on eye tracking data. They evaluated their model on natural appearance and effectiveness by comparing it with static and random saccades. The gaze model based on real eye behavior data outperformed the random and static gaze behavior.

Bente et al. (2007) examined the perception of social gaze in a virtual agent mediated communication. They measured the effect of varying durations of directed gaze with two studies. They developed a platform that is able to track the user's torso, arm, head, hand and eye movements through motion capture sensors, data gloves and an eye tracker. This system is able to fully track a human, display it through a virtual counterpart and to alter algorithmically specific modalities. For the two studies Bente et al. (2007) altered the gaze behavior only. The first study tested variations of gaze duration among female dyads. The variations on gaze were real, short and long gaze. They found that longer phases of directed gaze lead to a friendlier perception of the gaze behavior. The second study tested gaze duration variations among mixed gender dyads and Bente et al. (2007) could

confirm that females respond more sensitively to gaze variations of the male interlocutor.

Khullar and Badler (2001) propose a framework that generates attentive gaze behavior for virtual agents in a virtual environment. They introduce an intention list that drives the visual behavior of an agent. The intention list can be filled with sites and object of the virtual world. For example, during walking the virtual agents scans its environment and adds relevant items of the virtual world to the intention list. The system is further able to track and respond to moving objects like vehicles. In addition the framework is also able to track and monitor changing objects like traffic lights. This framework generates an autonomous gaze behavior for virtual agents dependent on the objects in a virtual world.

Gillies and Dodgson (2002) create a system for simulating attentive behavior to automatically generate gaze behavior for virtual agents. It allows to determine which objects in a virtual environment are possible to be perceived by the virtual agent. Further, the simulation of attention also determines where the character is looking to produce automatically appropriate gaze behavior. Gillies and Dodgson (2002) infer the parameters for the attentive model from observations of human behavior. The attention processing mainly consist of immediate and monitoring requests and a request can be defined through four parameters. The glance value defines whether the attention shift should be a short or long glance. The interval value gives the frequency for a monitoring object.

Gaze Models for Virtual Agents

Fukayama and colleagues propose a gaze behavior model for virtual characters based on amount and mean duration of gaze and averted gaze orientation Fukayama et al. (2002). They rated with two groups of attributes. One was named “friendliness” and was correlated in their study with attributes such as friendly, warm, sociable, tolerant, flexible, attentive and coordinative. The other was named “dominance” and correlated with assured, strong, successful, responsible and careful. Fukayama and colleagues found that a medium amount of gaze, a mean duration between 500 to 1000 ms conveys a “friendly” gaze behavior. The orientation of the gaze direction did not play a major role between the friendly and dominant gaze behavior, except a downward gaze was considered as less dominant.

In contrast to the previous two, Pelachaud and Bilvi (2003) propose a combined gaze behavior model based on statistical information of gaze patterns and communicative functions. They use a Bayesian Network to model the gaze behavior. The model integrates a speaking and listening mode and allows specifying the pa-

parameters for each mode independently. The current gaze is calculated dependent on the previous and current gaze target and the duration of the gaze. One parameter defines whether the virtual agent is looking at the interlocutor or is looking away, while a second parameter stores the previous state of the gaze direction.

While the previous works focus on the automatic generation of natural or neutral gaze behavior for virtual agents, [Lance and Marsella \(2007\)](#) examine how a virtual agent could generate emotional expressive gaze behavior. They use a 3-dimensional emotional model with pleasure, arousal and dominance (PAD) to describe the specific states for the gaze behavior. They use only arousal and dominance to describe the virtual agent's behavior in this work. In an evaluation [Lance and Marsella \(2007\)](#) could show that the encoded expressions could be recognized. In [Lance and Marsella \(2008\)](#) they developed a gaze model for emotional expression. They used the PAD model and a simple list of emotions to evaluate their gaze model with varying parameters. They found that a virtual character with either a raised head or a bowed body and/or fast movements appears more dominant. Low dominance was found for a bowed head and/or a neutral body posture without fast movements.

While all these approaches modify parameters, such as the orientation or the timing of changes in gaze, depending on whether the agent is speaking or listening, they do not track the users' gaze behaviors and can mainly be considered as a "one way" behavior.

2.3.2 Gaze during Non-Verbal Interaction

So far, embodied conversational agents have rarely been used as flirt partners. An exception includes the work by [Pan and Slater \(2007\)](#) who present a virtual party-like environment in which a female character called Christine approaches a male user and involves him in a conversation. The character shows her interest in the male user by smiles and head nods. She leans her upper body forward towards the user, looks at him and maintains eye contact with him. After some time, the character moves closer and formulates personal questions and statements. Using physiological measurements, [Pan and Slater \(2007\)](#) found out that the participants' level of arousal was correlated to compliments and intimate questions of the character. In addition, some of them indicated in a questionnaire that they had the feeling to have flirted with a real woman.

Unlike Pan and Slater, we do not make use of a full-body character, we just show the upper body of the agent in order to make sure that the users are able to perceive the subtle signals in the agent's face. Furthermore, Pan and Slater's agent is

not fully responsive since it is not able to recognize the user's gaze. Finally, the courtship behaviors of our agent are much more subtle concentrating on gaze behaviors and light smiles. We are less interested in the simulation of flirting per se, but rather aim at investigating to what extent courtship behaviors may contribute to the creation of instant rapport.

Gratch et al. (2006) developed a so-called rapport agent which acts as a silent listener. The agent tries to create rapport by providing rapid non-verbal feedback, expressed by head nods, posture shifts and gaze behaviors, through a shallow real-time analysis of the human's voice, head motion and body posture. An empirical study revealed that the agent increased speaker fluency and engagement.

The robotic penguin developed by Sidner et al. (2004) is able to track the face of the conversational partner and adjusts its gaze towards him or her. Even though the set of communicative gestures was strongly limited, an empirical study revealed that users indeed seem to be sensitive to a robot's conversational gestures and establish mutual gaze with it.

While most research focuses on how to create rapport in short-term interactions, Cassell and Bickmore (2003) investigate how to establish and maintain a relationship between a user and an agent over a series of conversations. They performed a series of experiments with a virtual character acting as a real-estate agent which revealed that the character's use of social language had an important impact on the creation of rapport.

Most studies analyzing the user's non-verbal feedback behavior make use of head trackers. They are able to roughly assess in which direction the user is looking, but do not have more detailed information on the user's gaze direction. One of the earliest work of using eye trackers for agent-based human interaction comes from Starker and Bolt (1990). They adapt "The Little Prince" to the users' current interest in a virtual scene that shows one planet from the story by Antoine de Saint-Exupéry. Dependent on the duration and focus of the user's gaze further details of the scene are described via a text-to-speech system. Another exception includes the work by Eichner et al. (2007) who made use of an eye tracker. In an experiment, they showed that agents that adapted the content of their presentation to a user's gaze were perceived as more natural and responsive than agents that did not have that capability.

The studies above show that embodied conversational agents are to a certain extent able to establish rapport with human conversational partners through appropriate verbal and non-verbal behaviors. Unlike the approaches described above, we focus on the first seconds of an encounter and investigate how to create a

friendly and natural atmosphere for human-agent communication by appropriate courtship behaviors of the agent. We make use of an eye tracker to monitor the user's gaze behaviors. A particular challenge of our work is to align and synchronize the user's gaze behaviors with the agent's courtship behaviors.

2.3.3 Gaze during Verbal Interaction

A number of studies informed by human-human conversation that investigate the role of gaze in human-agent communication provide evidence that natural gaze behaviors of an agent are not only more positively perceived, but elicit also more natural responses in human users (see, for example, Colburn et al., 2000; Garau et al., 2001; Lee et al., 2002; Vinayagamoorthy et al., 2004; Peters et al., 2005; Kipp and Gebhard, 2008).

Colburn et al. (2000) investigated whether natural gaze behaviors of an avatar elicit more natural gaze behaviors in users communicating with it. When an avatar was present, subjects spent more time looking at the screen. Even more attention was directed to the avatar when the agent relied on a gaze model that was informed by psychological studies on human-human conversation. Colburn and colleagues hypothesize that humans feel less shy when talking to a monitor than when talking to a real human. The effect occurred, however, only in the user-as-speaker condition which Colburn and colleagues attribute to the bad quality of the employed lip-sync mechanism. While Colburn and colleagues concentrate on the behavioral response to avatars employing an informed gaze model, Garau et al. (2001) as well as Lee et al. (2002) investigate the effect of informed gaze models on the perceived quality of communication by means of questionnaires. Both research teams observed a superiority of informed gaze behaviors over randomized gaze behaviors. A follow-up study by Vinayagamoorthy et al. (2004) focused on the correlation between visual realism and behavioral realism. They found that the model-based gaze model improved the quality of communication when a realistic avatar was used. For cartoonish avatars, no such effect was observed. While all these approaches modify parameters, such as the timing of changes in gaze, depending on whether the agent is speaking or listening, they do not track the users' gaze behaviors.

Step toe et al. (2008) used mobile eye trackers in order to drive the gaze behaviors of a user's avatar in a multiparty CAVE-based system. They found that gaze behaviors known from human-human communication also occurred in their 3D environment. For example, participants looked at the speaker when being asked a question or looked away when thinking of an appropriate response. The avatars

in their 3D environment just mimicked, however, the gaze behavior of the human users and did not generate gaze behaviors autonomously.

Rehm and André (2005) described an experiment where they investigated the user's level of attention in a multi-party scenario consisting of two human and one synthetic interlocutors. Their agent was not able to perceive the users. However, since the conversation followed a pre-defined sequence of turns, the agent knew whether the user to her left or to her right was speaking and could move her head into that direction.

Similar to Steptoe and colleagues, they found that certain gaze practices known from human-human conversation were followed. However, the users looked significantly more often to the agent when she was talking to them than when a human user was talking to them. The experiment left open whether this difference was caused by the novelty effect of the agent or by difficulties of the users to understand the agent.

Many systems investigating interactive models of visual attention make use of head trackers. They are able to roughly assess in which direction the user is looking, but do not have more detailed information on the user's gaze direction. Another application using an virtual agent is the MACK system (Nakano et al., 2003). The authors use a head tracker to determine a user's gaze in a direction giving task. The animated agent explains directions on a map and monitors the user's head. In this application, lack of negative feedback indicates successful grounding. If grounding fails, the agent will perform a repair action to help the user. Based on an analysis of human-human conversation, Sidner et al. (2004) developed a model of engagement for a conversational robot that was able to track the user's face and adjusted its gaze accordingly. Even though the set of communicative behaviors of the robot was strongly limited, an empirical study revealed that users indeed seem to be sensitive to a robot's conversational gestures and establish mutual gaze with it.

Another example is the FRED system by Vertegaal et al. (2001) that makes use of 3D animated facial agents in a multi-agent setting that are controlled by a conversational gaze model. The agents have the capability of noticing whether the user (or another agent) is looking at them. Together with the speech data they can determine whether they have to listen to someone else or whether they can talk. The focus of this work is the regulation of conversational flow in a multi-agent environment. That is the users' gaze in combination with their speech is used by the agents to determine whether to speak or to listen. Unlike Vertegaal and colleagues, we concentrate on mechanisms to establish mutual gaze and to respond to obtrusive staring behaviors in combination with turn taking. Eichner

et al. (2007) made use of an eye tracker to detect interest and attentiveness in a presentation. By means of an experiment, we showed that agents that adapted the content of their presentation to a user's gaze were perceived as more natural and responsive than agents that did not have this capability. The role of gaze as an important indicator for user attention and interest was also confirmed in a recent experiment by Nakano and Yamaoka (2009).

Chapter 3

Study I: Explicit vs. Implicit Gaze Interaction

This chapter examines how gaze based graphical user interfaces could be developed to support users during the interaction with a machine. In the first section we will discover what explicit (i.e. direct and deliberate control of an interface) gaze interaction with an interface could look like. It further takes into account how an explicit gaze controlled interface should take the nature of human eyes into account. The second experiment will investigate how an interface can be controlled by implicit interaction. In our case we will implement a recommender system that is based on the user's gaze.

3.1 Experiment I: Explicit Interaction

In this experiment we will investigate the usability of an eye controlled writing interface that matches the nature of human gaze, which always moves and is not immediately able to trigger the selection of a button. We classify writing into three categories (typing, gesturing, and continuous writing) and explain why continuous writing comes closest to the nature of human gaze. We propose Quikwriting, which was originally designed for handhelds, as a method for text input that meets the requirements of gaze controlled input best. We adapt its design for the usage with gaze. Based on the results of a first study, we formulate some guidelines for the design of future Quikwriting-based gaze controlled applications (Bee and André, 2008b).

Isokoski (2000) describes an adaption of Quikwriting for usage with gaze. Quikwriting was developed by Perlin (1998) as a new text input interface for stylus-based systems, i.e. handhelds or smartphones. The system is based on two input

concepts. First, with Quikwriting users must never lift their stylus from the surface. This approach perfectly matches the nature of human gaze. The eye is always gazing at something, e.g. the screen, and we cannot 'lift' our gaze from the screen unless we close our eyes. But then we can no longer see anything. Lifting, we better say triggering or switching, is not a natural human gaze behavior. Second, the user never may stop moving the stylus. Of course eyes stop often to fixate something in a scene, but these fixations normally just last around 150-300 ms. This is much shorter than the trigger time in a dwell-time system. As soon as the dwell time is equal to fixation time, everything users look at is selected.

The interface of Quikwriting is divided into eight equally sized sections around a central resting area. To write a letter, the user moves from the center to one of the outer sections, optionally to another adjacent section, and back to the center, which triggers the selection of the letter. Every section is linked to a group of letters. In general, the letters are arranged in such a way that frequent characters can be written faster. Thus training speeds up writing since users familiar with the arrangement would be able to find an intended letter faster than novice users. For instance, one section contains 'h', 'e', and 'c' in this order. To write an 'e', users simply move their stylus to this section and back to the resting area. If they want to write the 'h', they move to this section, after that to the adjacent left section and finally directly back to the central area. Isokoski (2000) never seemed to have implemented this system and thus results about its usability are not available.

3.1.1 Implementation of Eye Writing Applications

Our objective is to develop a new gaze-controlled dwell-time free system for continuous writing. We hope that such a system would come close to the nature of human gaze behavior. Based on earlier research, we will concentrate on an interface design which does not require learning any gestures by heart. Taking the Midas touch problem (Jacob, 1991) into account, our interface shall be comfortable for the eye. Therefore, distracting visual feedback should be handled with care. Taking these requirements into account, we decided to explore the potential of Quikwriting, which was designed for the usage with handhelds in the first place, for a gaze-controlled interface (see Figure 3.1). In addition, we will provide a comparison with a gaze-controlled keyboard-based system.

Gaze-Controlled Quikwriting

Two problems occur with the original design of Quikwriting when simply replacing the stylus by gaze. When interacting with a stylus, users first search for the

section that contains the letter to write and then they move the stylus there to select it. To avoid that the selection of a letter is unintentionally triggered by a visual search process, we decided not to display the letters within the original sections. Instead, we displayed them in the inner resting area, each group of letters close to its linked section. Further, we had to help users memorize which adjacent section is linked to which single letter, since users might already forget during the interaction which section they have to gaze at. This would be fatal for our system. Imagine that the user has selected a group of letters by moving his gaze out of the center to an outer section and now has to gaze at an adjacent section to trigger the selection of a letter. Let us assume that the user forgot whether the first or the second adjacent section is the correct one. Thus his gaze would move back again to the resting area. But this process already triggers the selection of a letter. To avoid this problem, we display each single letter within the section the user has to gaze at in order to write a letter after having selected a group of letters by moving his gaze out of the center (see Figure 3.1). Then he can look at the section for the letter to write and gaze back to the center.

Another problem occurs when a user wants to check what she or he already wrote. For instance, if we place the text area below the writing interface, users may gaze from the center at the text area and pass writing sensitive sections. Looking back to the center would result in the selection of a letter. To avoid this kind of unintended writing actions, we check whether the user is looking at the text field and disable any writing function until the gaze is back in the central area.

Gaze-Controlled Keyboard-Based System as a Baseline

The performance of writing systems differs a lot in the literature. Some authors measure in characters per minute and others measure in words per minute. Sometimes, the authors do not even describe how many characters make up their words. Most literature considers for European languages a sequence of five characters including spaces and punctuation marks for one word. MacKenzie (2003) describe in detail how to measure writing speed in theory and practice for various kinds of interaction devices. Not only the measurement methods differ frequently, also the writing speed of gaze-controlled keyboard-based systems, for instance, ranges from about 5 wpm (Wobbrock et al., 2008) to about 11 wpm (Urbina and Huckauf, 2007) for novice users. Majaranta et al. (2004) report that even among subjects the speed for gaze-controlled writing varies from 7 wpm to over 14 wpm. The huge variance could come from the different eye tracking systems and their accuracy, tracking capabilities and delay. Further, it is important to know if only correctly written letters or all written letters are taken into account. Including

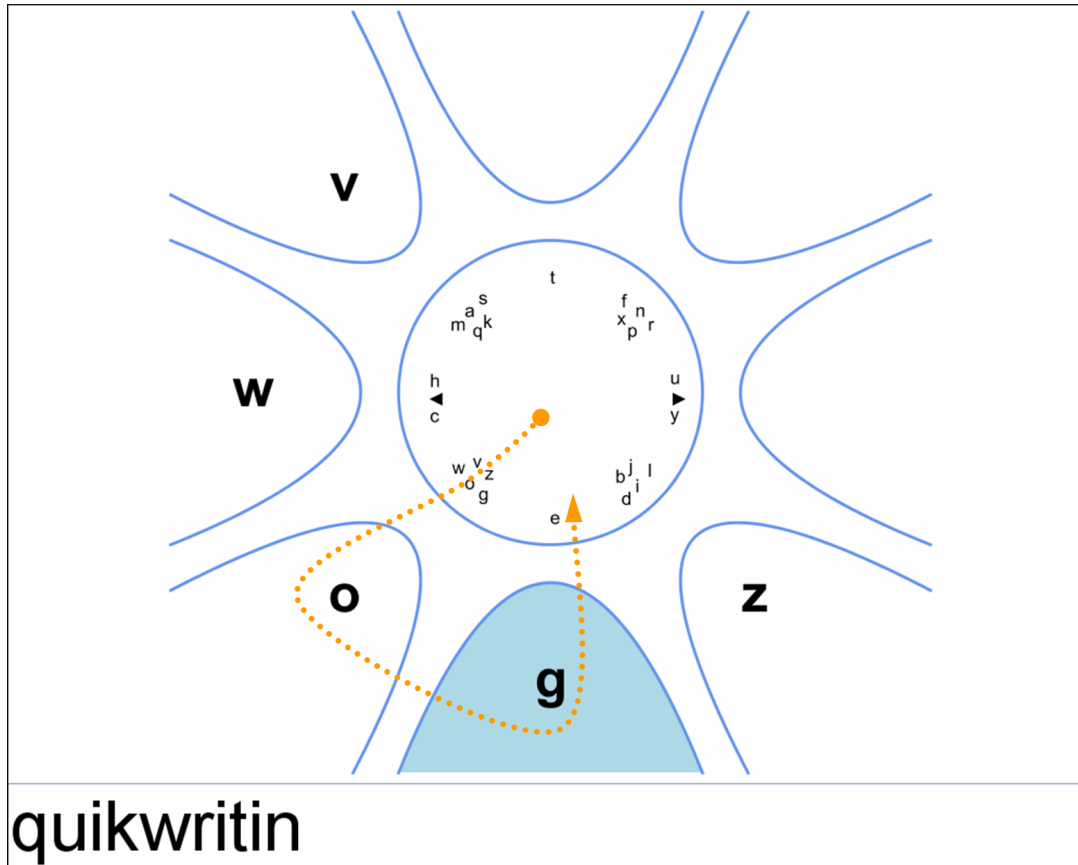


Figure 3.1: Adapted interface of Quikwriting for use with gaze. The dotted line indicates the gaze path to write the letter 'g'. The user currently looks at 'g'. The light shaded background follows the gaze and indicates which section the user is looking at.

wrongly written letters in our analysis would falsify our results, as they might have been written randomly and unwillingly.

This all makes it difficult for us to compare the performance of our newly developed application with results in the literature. As among systems without word completion, keyboard-based systems are the fastest way to write with gaze control, we decided to implement such a system. This gives us a trustier way to compare writing speeds. Our implementation of a keyboard-based system used a dwell time of 750 ms as our study only included novice users that never used an eye tracking system before. The system's response during the writing process was limited to visual feedback. Users were allowed to interrupt looking at a key to trigger it. Every key has its own dwell time buffer. As soon as a dwell time of one letter exceeds, all other buffers are reset. This gives users the freedom to look at the already written text while writing a letter. Also for eye tracking systems with lower accuracy, users won't get easily annoyed if the gaze leaves the key for

a glance and the dwell time is reset. While looking at one key, the background color changes and slowly fills the button from inside. The letters were arranged in alphabetical order. Additionally we had three command buttons for space, delete and enter.

3.1.2 Experiment

To investigate whether the new interface is usable and whether the writing speed can compete with a gaze-controlled keyboard-based interface, we conducted a study with 3 subjects.

We used the iView X RED eye tracking system from SensoMotoric Instruments (SMI), which is contact-free. The eye tracker operates with a sampling rate of 50 Hz and is used in combination with a 19-inch screen with a resolution of 1280×1024 . When a subject is placed in front of the eye tracking system, the system automatically recognises the position of the head and gives hints about its best position. While the subjects' gazes are tracked, they are allowed to move their head freely to a certain extent.

After the subjects were placed in front of the eye tracker we gave them a short introduction about the eye tracking system. We explained to them how they would use their gaze to write. Before we started the study, we gave them about 5 minutes to get used to the applications, as our subjects never used an eye tracker before. This was to ensure that the subjects were able to work with our eye tracking system and the tracking accuracy was high enough. We prepared 30 short sentences on small index cards, which were shown and read to them before they started to write a sentence. Every subject had to use the adapted Quikwriting system and the keyboard-based system. Per application they had to write 10 sentences. The sentences were selected randomly per subject. We showed the cards to the subjects to avoid misspelling which would have a side effect on the analyzed error rate. The applications logged the users' writing interaction into a file with each interaction and a time stamp.

3.1.3 Results

We analysed the log files and removed all unwillingly and wrongly written letters. This was necessary as sometimes the users wrote letters although they did not intend to do so. And as we were only interested in the writing speed of correctly written letters the wrong ones were removed. Removing the wrongly written letters normally worsens the writing speed. Writing unwillingly letters often

occur randomly and unexpectedly and therefore the selection of such letters takes a shorter time. We observed (see Table 3.1) that users were able to write with our adapted Quikwriting 5 wpm. The same subjects achieved a writing speed of about 8 wpm with the keyboard-based interface.

Table 3.1: Comparison of writing speeds in wpm (words per minute)

	eyeKeyboard	Quikwriting
avg	7.8	5.0
var	0.02	0.3

For all users the keyboard-based system was easier to use, but more exhausting. This was surprising to us since the users were familiar with keyboards, while the adapted Quikwriting was new to them and its usage had to be learned. The reason for the keyboard-based system to be exhausting is the Midas Touch problem. The users always feared that something happened, when they looked somewhere. Some asked if there is a place where nothing happens and they can rest their eyes. In the adapted Quikwriting we have automatically a resting area in the middle of the interface. On the other side the speed of the keyboard-based system could be increased by reducing our current dwell time of 750 ms. For instance, with a dwell time of 500 ms the writing speed would increase to about 9 wpm. An experienced (i.e. training of 5 days with 15 minutes each day) user could increase writing speed of Quikwriting up to 9.5 wpm, which comes close to a keyboard-based layout. Quikwriting will improve more as the letters are arranged dependent on their probability, and thus requires some time for training.

The error rate of the Quikwriting based system was much higher than the error rate of the keyboard based system. After analysing the log data we found that errors mainly occurred when the user checked what he or she already had already written. We further found that the blocking of any interaction after the users looked at the text field with the already written text did not work in a satisfactory manner. Obviously, users did not focus on the written text, as a half glance already sufficed for a human eye to recognise what was written. Therefore, we intend to change the layout of the interface. In particular, we plan to place the text field in a next step to the middle (see Figure 3.2). This will match the layout of Quikwriting as the gaze starts and stops in the middle resting area. Another point is the arrangement of the letters. We kept the layout of Quikwriting as it was originally designed by (Perlin, 1998). This includes the probability distribution of the characters in the English language. The first six most probable letters in English are E, T, A, O, I, and N, whereas in German they are E, N, I, S, R, and A.

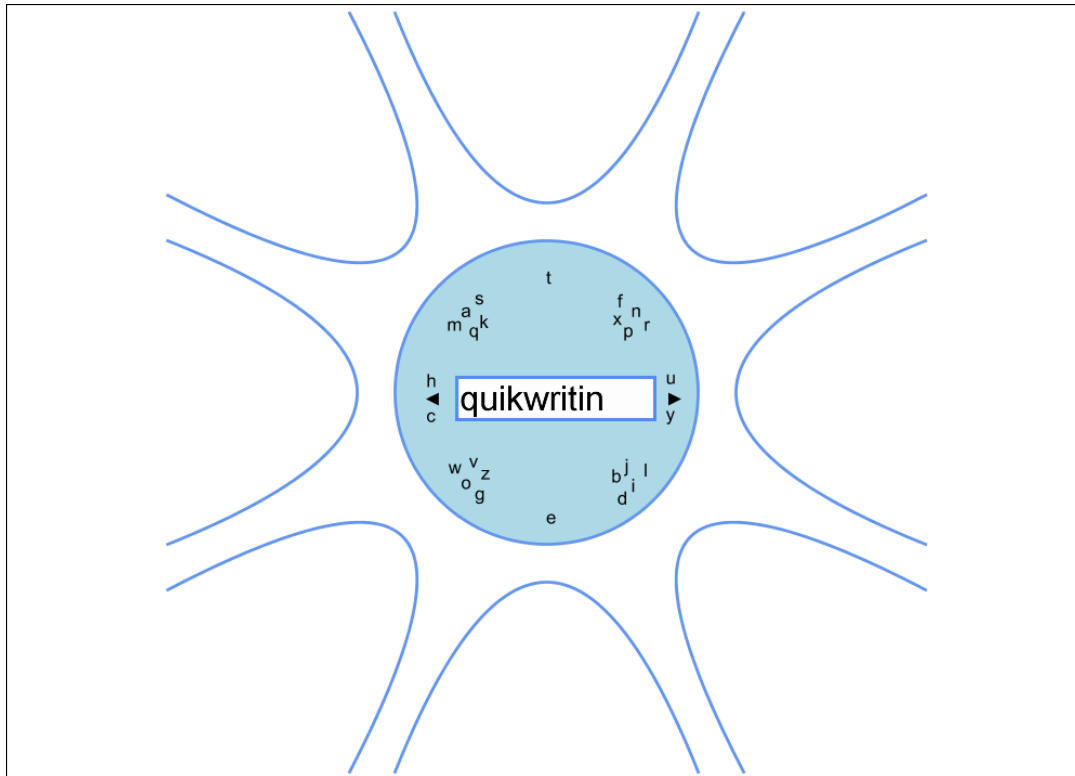


Figure 3.2: In a new design of the adapted Quikwriting, we will place the text field in the resting area. This will enable users to check what they have written without moving their eyes to any place outside the control interface.

The keyboard-based system could be improved by providing audio feedback to the users after a letter was written. Indeed, subjects already looked at the next letter although the current one was not yet written. Also the dwell time should be adjusted to the users, as for some of them staring for 750 ms was too long.

Comparison With Other Input Methods. The speed of handwriting is about 30 wpm. An average typist writes with 50 to 70 wpm on a computer keyboard. Of course gaze controlled writing systems cannot compete with these input methods, but users would need their hands. An input method that needs hands as well are game controllers. Normally the controller is used to select the letters on a keyboard-based interface displayed on the screen, similar to the one we described in section 2.1.1. Költringer et al. (2007) conducted a study with common game controllers. They found that the writing speed of novice users is about 8 wpm using such devices. Experienced users after 15 sessions were able to write 12 wpm. The speed of this input method is comparable to that of gaze controlled systems. Since text input becomes more and more important for gaming consoles, gaze based input methods will become interesting as soon as webcam-based eye

tracking systems become more accurate and reliable.

3.2 Experiment II: Implicit Interaction

In contrast to the previous experiment, the following experiment illuminates the advantages and disadvantages of implicit gaze interaction. We chose an example (see also Bee et al. (2006a) and Bee et al. (2006b)) that predicts (visual) preference decisions of users in real-time. This kind of system would recognize a user's choice of a particular visually presented stimulus in the presence of other stimuli, and respond accordingly. Our system, called AutoSelect, may automatically detect a user's visual preference solely based on eye movement data in a two-alternative forced choice (2AFC) setting. This experiment was conducted in collaboration with the National Institute of Informatics, Japan at the lab of Prof. Dr. Helmut Prendinger.

We believe that visual attention based interactive technology is of high relevance to various applications, including e-learning, future interfaces, as well as devices for handicapped people. In fact, many decisions of our daily life can be reduced to choices between several items, and cannot be easily explained in terms of overt reasoning on premises. In a restaurant, for instance, we choose between different types of dishes. Unless price or dietary considerations are of primary importance, our decision for a particular dish might be based on our taste, our expectation of a specific (eating) experience, or even our current mood.

The analysis of gaze patterns may provide an effective means to unveil non-conscious preference decisions of people. This experiment describes AutoSelect, our system that exploits the gaze 'cascade effect' and a recently conducted pilot study.

3.2.1 Gaze Cascade Effect

When presenting pairs of human faces to subjects and giving the instruction to decide on their attractiveness, Shimojo et al. (2003) observed a phenomenon they called gaze 'cascade effect'. This phenomenon involves the gradual gaze shift toward the face that was eventually chosen (as more attractive), while gaze bias was initially distributed evenly between the two presented faces. The results of the two-alternative forced choice (2AFC) task used in their study demonstrated a progressive bias in subjects' gaze toward the chosen stimulus (preference formation), which was measured by the gaze time spent on the selected stimulus. However, the strong correlation between choice and gaze duration occurred only in the last

one and half seconds before the decision was made. A finding that Shimojo et al. (2003) declared as surprising relates to the result that a larger cascade effect was found in the ‘difficult’ task, where the comparison between the attractiveness of faces was difficult, while intuitively subjects were expected to more evenly distribute their gaze between stimuli in this case in order to compare stimuli in as much detail as possible. The result was explained by a theory claiming that gaze would significantly contribute to decision-making when cognitive bias is weak. The importance of this result for our research derives from the fact that a large number daily choices, e.g. regarding consumer products, are also deficient of a strong cognitive bias, and hence contributes to the importance of investigating non-conscious human decisions.

3.2.2 Study

A system that is able to automatically detect users’ choices seems to break new ground. We therefore conducted an exploratory study using the AutoSelect system. Our first application is an automatic necktie selector, where subjects are shown a pair of ties and the AutoSelect system tries to detect the preferred tie. Subjects were given no instruction other than having to choose a tie for themselves or their friend for a graduation party.

We used faceLAB v4¹, a non-contact vision-based system with a sampling rate of 60 Hz. We implemented an algorithm based on the findings of Shimojo et al. (2003), which detects visual preference in real-time.

Eight subjects (4 female, 4 male), all students or researchers from the NII in Tokyo, participated in our study. Subjects entered the experimental room individually and were provided written instructions about their task. Subjects were seated in front of a 20.1 inch display with attached infrared lights and their head and eyes were calibrated. This procedure had to be performed for each individual once, and took approximately 5 minutes. A session was initialized by subjects pressing a ‘start’ button in a web page based interface (see Figure 3.3).

The following procedure was then iterated for 62 pairs of ties. First, a center located ‘dot’ was shown on the screen for 2.5 s in order to eliminate any initial gaze bias. Next, a pair of ties was presented, located to the left and to the right on the screen. In order to guarantee that subjects actually compare the ties, automatic selection was suppressed within the first 2.5 s. This value was based on the empirically determined decision time of 4 s (Shimojo et al., 2003). After the system decision, the selected tie was presented and subjects were asked to indicate

¹<http://www.seeingmachines.com/product/facelab/>

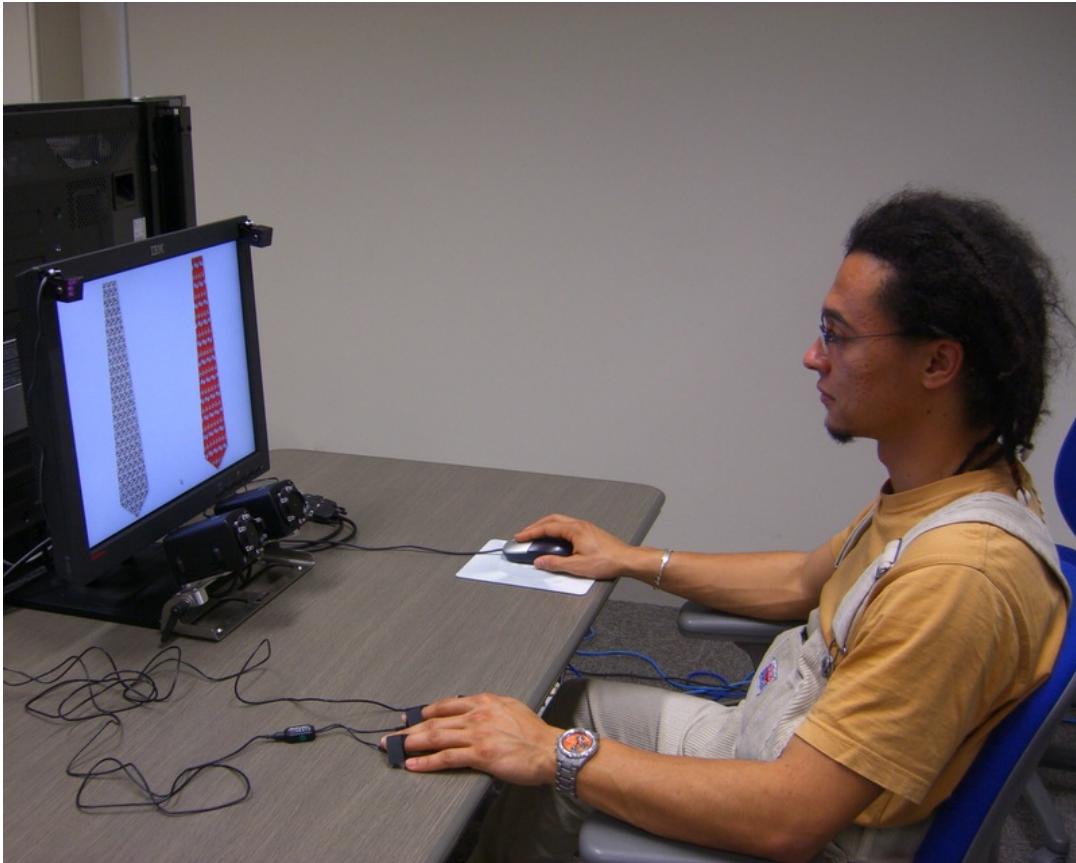


Figure 3.3: The user sits in front of a 19 screen and an eye tracker. The system automatically selects the preferred tie dependent on the user's gaze behavior.

whether the system choice is correct by clicking a 'yes' or 'no' button. Then the next iteration started with the initial view of a center dot. One initial set of 32 tie pairs was prepared, and the chosen ties were put back into the tie pool, which was used to create the subsequent set of 16 pairs, and so on. Eventually, subjects were shown a single pair of ties they presumably liked best. Hence, subjects were exposed to 63 pairs and performed 62 decisions in total. In the initial set of tie pairs, two partitions were created with 13 pairs each. One partition contained pairs of 'different' type ties, i.e. formal (decent) vs. 'entertainment' (adventurous) style ties, whereas the other partition contained 'similar' type ties that differed only in color or had a slightly different pattern but the same color. The motivation of this grouping was to investigate differences in subjects' decision behavior for presumably 'easy' vs. 'hard' decisions. All sessions were logged and lasted for about 10 minutes.

3.2.3 Results

The primary result concerns the classification accuracy of the AutoSelect system. In our study, the system was able to detect subjects' choices correctly in 81 % of the cases. The worst recognition rate was 68 %. Given a chance level of 50 %, the system performed very well. (One subject was excluded from the analysis because of distorted values due to starting a conversation during the experiment.) We wanted to investigate the users' interactive experience with a running system, which can reveal e.g. issues related to the latency between user decision and system decision. Informal comments on the system indeed indicated that subjects were surprised about the system's reliability to timely identify which tie they liked more. Some of the misclassifications were related to a design problem, i.e. when subjects moved their face out of the camera range. The next version of AutoSelect should alert subjects in those situations. Furthermore we were particularly interested in results comparable to the 'difficult' vs. 'easy' choice finding reported in (Shimojo et al., 2003).

We hence compared recognition rates and decision times for 'different' vs. 'similar' tie pairs. Recognition rates were 75 % (different ties) and 81 % (similar ties); decision times were 6.8 s (different ties) and 7.65 s (similar ties) In line with (Shimojo et al., 2003), the decision time for different ties was significantly longer than for similar ties ($t(180) = -1.66; p < 0.05$). A one-tailed t-test assuming unequal variances was used in our analysis. This result supports the hypothesis that a choice between unlike items relies on (time consuming) cognitive processing, whereas similar items might be chosen based on non-conscious ('intuitive') preference. We also note that the system calculated the choice between similar ties more accurately.

3.3 Conclusion

Explicit gaze interaction has the potential of becoming a new form of interaction in human-computer interfaces. Currently the interest in such interaction systems mainly comes from physically handicapped people that, for instance, cannot keep their hands calm or move them at all.

We developed a new writing system for gaze controlled interaction. Our very first prototype can easily compete with gaze-controlled keyboard-based systems. As we were testing the system for the German language, we expect an improvement after we place the letters according to the occurrence probability of German letters.

And with moving the text field from the bottom to the center, a more continuous flow of writing should become possible and increase writing speed.

Based on the results of the first experiment, we formulated some guidelines for the design of future Quikwriting-based gaze control. Quikwriting was originally designed for the usage with stylus-based input devices. The underlining principles of Quikwriting, (1) always move and (2) never lift the stylus, perfectly match the nature of human's gaze and should be considered in future designs for gaze interaction. We were able to show that such a system can compete with common writing systems without word completion for gaze, such as GazeTalk or pEYWrite.

Also the results achieved in the second experiment show great potential for future gaze-based implicit interfaces. For example, a recommender system could simply use the gaze cascade effect to detect users' preference.

Chapter 4

Framework for Gaze-Based Human-Agent Interaction

To examine gaze based human-agent interaction we need a framework for virtual agents that is able to control head, face and eyes of a virtual character. The framework is named Horde3D GameEngine and this chapter explains the most relevant components of the Horde3D GameEngine for controlling a virtual agent, i.e. facial expression control, lip synchronization, gaze control and object detection, that were used to control a virtual agent. The second part of this chapter deals with the usage of these components and the evaluation how facial expression control can be simplified.

4.1 Horde3D GameEngine

The Horde3D GameEngine developed by [Augsburg University](http://horde3d.org) (2007) extends the Horde3D¹ graphics engine. The Horde3D graphics engine is a state of the art shader-based rendering engine for animated 3D computer graphics. Thus, it allows rendering high quality scenes. The simple integration of the graphic shaders into the rendering pipeline allows easy application of different material shaders. For example, the virtual character Alfred (see Section 4.2.2) was originally developed with subsurface scattering for the skin and a reflection shader for the eyes (see [Schulz, 2008](#), for more details). The Horde3D graphics engine, the Horde3D GameEngine and the virtual character Alfred are all open source and can be used for own applications.

The idea behind the Horde3D GameEngine is to extend Horde3D with often used functions like methods to handle animations, to play sounds or to detect collisions

¹<http://horde3d.org>

of objects within the virtual scene. The Horde3D GameEngine provides a component based game engine approach. All objects within a game application are represented as an entity in a game world. Each entity can have several unique components. The components are implemented as plugins that can be loaded dynamically. This lets one easily add or remove functionalities from single 3D objects, which can be configured through an XML file. They interact with each other using game event messages. The core provides the communication system, the plugin management and the entity class.

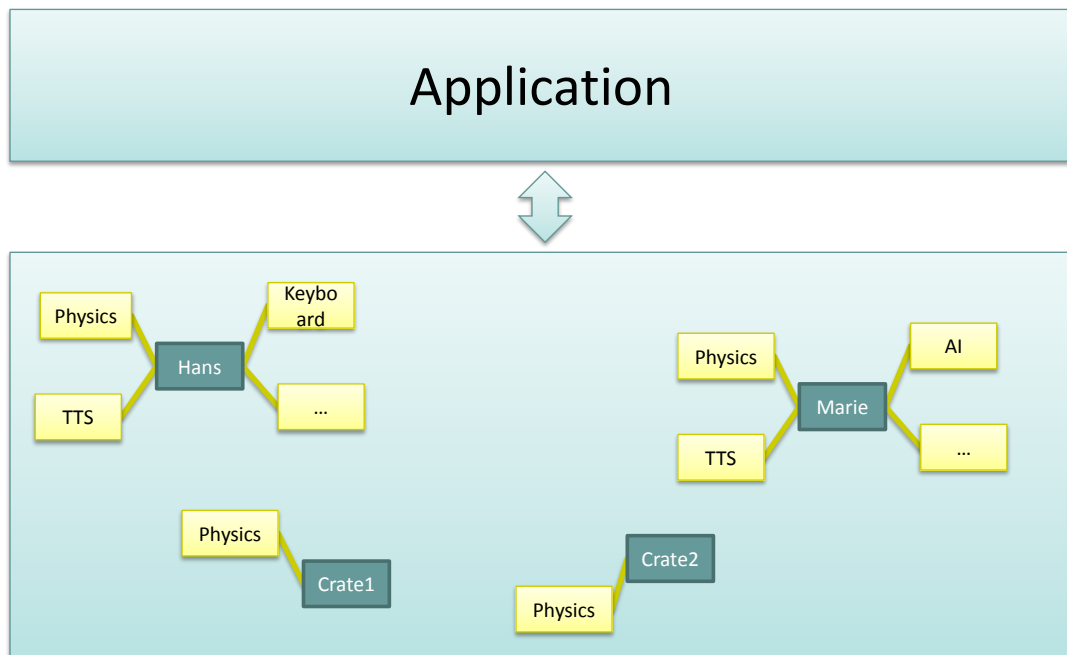


Figure 4.1: The game application (top) interacts with the game world (bottom) containing several entities (e.g. Marie or Crate1) each having one or more components (e.g. Physics or TTS).

In Figure 4.1 you can see the game application (top) which interacts with the game world (bottom) containing several entities (e.g. Marie or Crate1) each having one or more components (e.g. Physics or TTS). For instance, the Keyboard-component is attached to Hans, which simply allows the translation of this entity with the keyboard arrows. The functionality of translating and keyboard mapping is all hidden in the keyboard component.

For the gaze based interaction with virtual characters, we had to create a component for controlling the facial expressions of the virtual character, a component that controls the lip synchronization during speaking, a component that is able to control the virtual character's head and eye orientation and a component that is able to detect the objects in the virtual world that are currently seen by a real user.

4.1.1 Facial Expressions

Our facial expression component bases on the Facial Action Coding System (FACS) by Ekman and Friesen (1975). FACS was originally designed to analyze natural facial expressions, but it turned out to be usable as a standard for production purposes too. That is why FACS based coding systems are used with the generation of facial expressions displayed by virtual characters, like Gollum in the movie trilogy *The Lord of the Rings* (2001-2003), King Kong in Peter Jackson's *King Kong* (2005), the characters in *Monster House* (2006) (Sagar, 2006) or all the Na'vis in the movie *Avatar* (2009). But the usage of FACS is not only limited to virtual characters and facial expressions in movies. The gaming industry with *Half-Life 2* by Valve also utilizes the FACS system for the first time to produce the facial expressions of their game characters. One of the most recent games is *Harry Potter and the Deathly Hallows* by Electronic Arts.

FACS defines 32 so called "Action Units" (AU) which are motivated through the human facial muscle system (e.g. inner brow raiser, upper lid raiser, or lip corner depressor). The action units describe the movement of a facial part controlled by one or several muscles. FACS consists of 32 action units and additionally of 26 action descriptors, which describe more complex movements outside the mimic muscles, e.g. the rotation of the head or the eyes.

To control the facial expression of a virtual character with FACS, we use morph targets (also known as blend shapes; see for example (Spencer-Smith et al., 2001) or (Frydrych et al., 2003)). Morph targets describe the translation of a set of vertices to a defined new position in the 3D space. In our implementation, each morph target represents one of the action units. As not all action units are necessary – some of the action units overlap and are not needed for generation – we could limit them to 23.

Another system, mostly used in academia, to generate facial expressions is the MPEG-4 standard (Pasquariello and Pelachaud, 2001), (Balci, 2004). It defines 66 facial action parameters (FAP) which control specific regions of the face (e.g. shift tongue tip, raise left middle eyebrow, stretch left corner lip). The basic principles of controlling facial expressions with the MPEG-4 standard and FACS are the same (Pandzic and Forchheimer, 2003).

We chose the FACS-based approach for our facial animation system, because of the availability of the *Facial Expression Repertoire* (2008), which maps over 150 emotional expressions to the action units of FACS. Not only does it explain in detail which action unit must be activated for certain facial expressions, it further provides a rich dataset of videos which show how the action units ought to be

designed. This rich database simply allows automatically generating a large set of facial expressions.

Morph Target Animation The Horde3D graphics engine offers the functionality to set the morph targets of any virtual shape. A morph target, which comes with the 3D model, can be set in Horde3D between 0 and 1 and all involved vertices in between are linearly blended. To animate the morph targets two parameters need to be considered: duration and target weight. For instance, if you want to change a neutral mouth to the maximum of a smiling mouth within 0.5 seconds, you need to linearly set the morph target weights from 0 to 1 within 0.5 seconds. To guarantee a smooth animation and the maximum after 0.5 seconds, you need to link the weight steps with the frame rate.

To enable the morph target animation, you first need to enable it in the XML file that describes the scene within the related entity:

```
<MorphTargetAnimation />
```

This enables the responsible component for controlling the morph target animation of any attached virtual object. To animate morph targets now, you simply need to know the name of the morph target and set the weight and the duration of the animation.

```
MorphTargetAnimation morphTargetAnimData(  
    morph_target_name , weight , duration );
```

FACS Component The Morph Target Animation component simply controls the morph targets of any virtual shape. The FACS component extends this feature to control facial expressions of any virtual head. Some requirements, though, need to be fulfilled for such a virtual head. First, the 3D model needs to have specific morph targets that come along with the Facial Action Coding System (FACS). This requires that the single Action Units needed for the facial expressions are available as morph targets within the virtual model. Further, the morph targets need to follow a specific naming. The single action units need to be numbered according to the numbers given by the Facial Action Coding System and are named internally like this AU_XX.

It is also necessary to predefine the facial expressions. The FACS component loads an XML file that contains facial expressions divided into the single action units. The (Facial Expression Repertoire, 2008), for instance, contains over 150 emotional expressions. This database can be used to easily define a large number of facial expressions.

```
<FacialExpression name="joy">
  <AU id="6" value="1.0" />
  <AU id="12" value="0.8" />
  <AU id="20" value="0.2" />
  <AU id="27" value="0.1" />
</FacialExpression>
```

For instance, to create a joy you need to activate four action units (i.e. 6 = *cheek raiser*, 12 = *lip corner puller*, 20 = *lip stretcher* and 27 = *mouth stretch*). The FACS component offers a function to set and animate the facial expression.

```
FACSControlComponent::setFacialExpression(
    expression_name, weight, duration )
```

The parameter `expression_name`, which selects one of the predefined facial expressions from the XML file. The `weight` gives the target intensity of the facial expression, which is a value between 0 and 1. The `duration` sets the time of the facial animation.

4.1.2 Lip Synchronization

The lip synchronization maps a phoneme of speech to a viseme. While a phoneme is a small unit of sound of an utterance, a viseme is the corresponding visual representation of the lips. For a talking head it is necessary to map the spoken utterances to the corresponding visemes in real-time. The lip synchronization for a virtual talking head within the Horde3D GameEngine offers two ways to let the virtual head speak. It can either talk via Text-to-Speech (TTS) or via prerecorded audio files with speech. For both approaches the Microsoft Speech API is used.

The Text-to-Speech (TTS) based approach for lip synchronization needs a TTS voice like it is, for example, offered by Loquendo². The TTS system delivers the spoken phonemes in real-time. These phonemes are caught through a callback method by the TTS component of the Horde3D GameEngine directly from the Microsoft Speech API and are mapped to the corresponding visemes.

The second option to let the virtual head talk is to record speech in advance. A tool included in the Horde3D GameEngine extracts the phonemes with the corresponding timings via the speech recognition functionality of the Microsoft Speech API. The phonemes and timings are then saved in XML format, which can easily be read from the Sound component of the game engine.

²<http://www.loquendo.com/>

The lip animation itself is also using the FACS component. The visemes can be, as well as the facial expressions, defined through the Facial Action Coding System. The FACS provides 18 action units corresponding to the muscles around the mouth and thus can be used to model the visemes. A viseme definition looks quite similar to a facial expression definition.

```
<FacialExpression name="r">
  <AU id="10" value="0.2"/>
  <AU id="20" value="0.04"/>
  <AU id="25" value="0.83"/>
  <AU id="27" value="0.19"/>
</FacialExpression>
```

For instance, to create the viseme *r* you need to activate four action units (i.e. 10 = *upper lip raiser*, 20 = *lip stretcher*, 25 = *lips part* and 27 = *mouth stretch*). The FACS component offers a function to set and animate the facial expression.

The system blends the current viseme with the past viseme. With the TTS-based approach the blending starts as soon as the TTS delivers the phoneme for the current spoken sound. The prerecorded approach takes the phonemes directly from the annotated utterance. This allows for a smooth animation of the lips during speaking.

4.1.3 Inverse Kinematics for Gaze

The Horde3D GameEngine includes an inverse kinematics component. This component uses the Cyclic Coordinate Descent (CCD) method to compute a solution for the given chain. The kinematic chain is formed out of the scene graph nodes of a model and usually represents a body region of an agent. The parameters for the inverse kinematics component are defined in an XML file.

```
<IK file="model_scene" l_eye="left_eye_name" r_eye="
  right_eye_name" head="head_name" />
```

The `model_scene` sets the file name of the scene file the model is defined in. It needs to contain the joint hierarchy of the model. The `left_eye_name` defines the joint (node) name of the left eye as it is defined in the model scene file, the `right_eye_name` defines the joint name of the right eye and the `head_name` defines the joint name of the head. The gaze functionality of the inverse kinematics component controls the eyes and the head of a virtual character by simply giving the gaze target point in the 3D space.

```
int IK_gaze( unsigned int entityWorldID, float targetX,
            float targetY, float targetZ, bool moveLEye, bool
            moveREye, bool moveHead, int head_pitch )
```

entityWorldID: The entity we want to use the function on

targetX: The value on the x-axis of the requested position

targetY: The value on the y-axis of the requested position

targetZ: The value on the z-axis of the requested position

moveLEye: Flag for moving the left eye

moveREye: Flag for moving the right eye

moveHead: Flag for moving the head

headPitch: A value from (-10, 10) representing the pitch the head will sustain during a gaze action (where -10 means a higher angle for the “pointing joint” therefor a “less arrogant” gaze and +10 the opposite)

return status code of type IK_GazeResult

The inverse kinematics component is not limited to control head and eyes of a virtual character. It also is able to solve chains for any node chain of a entity. For instance, it will be no problem to control hands and arms of a virtual character with the same component (Damian, 2011).

4.1.4 3D Object Detection

The 3D object detection component allowing detect predefined objects within the 3D world through screen coordinates. This enabled the Horde3D GameEngine to select or pick 3D objects with the cursor of a computer mouse or an eye tracker for example (see Figure 4.2).

In contrast to for example Pfeiffer and Latoschik (2004), we follow a simple approach for object detection in 3D virtual worlds. We use simple linked 3D hull objects to detect a user’s interest. The linked object can be a square or sphere shape, which is set with `object_type`. It can be scaled (`sx`, `sy`, `sz`) and translated (`tx`, `ty`, `tz`) relative to the position of the linked node (`node_name`).



Figure 4.2: Visualization of the picking objects for detecting the eyes and the head of Alfred.

```
<Reference tx="x" ty="y" tz="z" sx="x" sy="y" sz="z" rx="
  x" ry="y" rz="z" sceneGraph="object_type" name="
  ref_name">
  <Attachment type="GameEngine" name="link_name">
    <LinkedObject name="node_name" />
  </Attachment>
</Reference>
```

With this component one can easily use an eye tracker which delivers the screen coordinates of a current user looking at a display to pick these objects. Thus, it allows us, for example, to detect when and how long the user is looking at the virtual characters head or eyes. This is an important feature to be able to establish mutual gaze between a user and virtual character, for example.

4.2 Experiment: Facial Expression Control for Virtual Characters

We describe the usage of the Horde3D GameEngine to generate and control facial expressions in the following section. Editing facial expressions of virtual characters is quite a complex task. The face is made up of many muscles, which are partly activated concurrently. Virtual faces with human expressiveness are usually designed with a limited amount of facial regulators. Such regulators are derived from the facial muscle parts that are concurrently activated. Common tools for editing such facial expressions use slider-based interfaces where only a single input at a time is possible. Novel input devices, such as gamepads or data gloves, which allow parallel editing, could not only speed up editing, but also simplify the composition of new facial expressions. We created a virtual face with 23 facial controls and connected it with a slider-based GUI, a gamepad, and a data glove. We first conducted a survey with professional graphics designers to find out how the latter two new input devices would be received in a commercial context. A second comparative study with 17 subjects was conducted to analyze the performance and quality of these two new input devices using subjective and objective measurements (Bee et al., 2009b).

4.2.1 Control of Facial Expressions

Virtual worlds, such as Second Life, Lively by Google, or World of Warcraft, provide a rich platform for embodied interaction between people all over the world through the internet. The social component of such platforms is a fundamental part of their success. When it comes to close interaction, facial emotional expressions play an important role as non-verbal behavior to underline the written words during a chat. Especially in game-based multi player platforms, such as World of Warcraft, where users conduct quests with 2-40 companions, expressing emotions, for instance, becomes essential after you succeed or fail to accomplish a cooperative quest.

While the visual capabilities for displaying expressions through virtual characters has advanced quickly in the last few years, the control of facial expressions still remains a challenge. Common tools to adjust facial expressions use slider-based graphical interfaces that allow users to edit one facial parameter after the other. However, facial expressions involve several facial muscles in parallel. As a consequence, new input devices which support intelligent parallel control should not

only speed up the editing of facial expressions, but also simplify the editing for inexperienced users.

It turned out that the control of facial expressions of a virtual character is quite complex. It takes several controls in parallel to generate facial expression. Thus, we first discuss related work on facial animation control systems. Slider interfaces bear the advantage that they are both easy and quick to implement. Furthermore, most users are familiar with this kind of interface. Nevertheless, there are some serious pitfalls to be considered. First of all, users may only manipulate one parameter at a time. Yet, the interplay of different parameters is crucial in generating high quality facial animations. As a consequence, users need to switch back and forth between different sliders to adjust the parameters for the desired facial expression. Furthermore, the use of sliders is hardly intuitive since there is no obvious mapping between the manipulation of a slider and the movement of the corresponding mimic muscle. In order to know what effect a particular slider achieves, the user needs to interpret the description of the slider correctly. Uncommon anatomic technical terms may further hinder the user's understanding.

Approaches to facial animation control based on the manipulation of images also offer alternative solutions, for example see (Sucontphunt et al., 2008). The so-called sketch-based interfaces as previously introduced by Chang and Jenkins (2006) or Nataneli and Faloutsos (2007) go a step further and generate facial expressions from sketches drawn by a user. Jacquemin (2007) developed a 3D interface of editing facial expressions. The tangible interface named Pogany maps a real model of a human face to a computer generated 3D face. Depending on which region is touched on the physical model, the virtual match is activated and enables one to compose a facial expression. Jacquemin could show that such a novel interface is easily accepted and engages users in a pleasant way.

While these interfaces need special mapping, including pattern recognition or even special hardware to control facial expression of virtual faces, Thalmann (1993) analyzed a variety of more common hardware devices for animation control, including position and orientation trackers, data gloves, data suits, 6D-devices and midi keyboards. Particular emphasis was given to data gloves and midi keyboards as promising control devices for facial animation. The computer game "Indigo Prophecy" by Quantic Dream provides evidence of the practical use of data gloves. The facial expressions of this game were produced by translating the finger bends of the gloves an animator was wearing into the corresponding morph target animation parameters. The facial animations included emotional expressions as well as lip syncing.

The aforementioned mapping approaches use a direct mapping to facial muscles

or regions to control the facial expression of a virtual agent. They could be described as *direct* mapping. Another, *indirect*, approach to map emotions to facial expressions is to use a descriptive or a model representation of emotions. Ruttkay et al. (2003) developed EmotionDisc, where discrete emotions are arranged in a circular way. The distance from the center of the disc is always equivalent to the intensity of the current emotion dependent on the current angle. Albrecht et al. (2005) describe the usage of an emotion dimension model recommended by Cowie et al. (2000) to control the facial expressions of a virtual character. This model describes emotions in an activation-evaluation space. Depending on spatial position, the respective facial expression is displayed (e.g. the center displays neutral, the upper right area display happy or excited, ...). Courgeon et al. (2008) use a 3D model to describe emotions. They place a discrete emotion on every corner of a cube. Users control the 3D representation of it with a joystick and, depending on the position in the 3D space, an appropriate blended facial expression will be generated.

This way of controlling emotional facial expressions does not require the understanding of how to design or model facial expression. Thus, it makes it easily usable for inexperienced users.

4.2.2 FACS-based Facial Expression Generation

“Alfred” (see Figure 4.3) is a butler-like character used to display facial expressions. Alfred’s facial animations are based on the Facial Action Coding System (FACS) by (Ekman and Friesen, 1975) (see Section 4.1.1). To implement FACS in Alfred, morph targets (also known as blend shapes) were used (see e.g. (Spencer-Smith et al., 2001) or (Frydrych et al., 2003)). They describe the translation of a set of vertices to a defined new position in the 3D space. In our implementation, each morph target represents one of the action units.

Alfred’s mesh has a resolution of about 21.000 triangles. For displaying more detailed wrinkles in the face, normal maps baked from a high-resolution mesh are used (Oat, 2007). The morph targets for the action units are modeled using the actor’s templates from the FER. For rendering the character and its animations the Horde3D graphics engine developed by (Augsburg University, 2007) is used.

To control Alfred’s facial expressions (i.e. action units), we use the UDP network protocol. This allows us to easily connect new interfaces to control the virtual face. Any controller can send the desired expression in terms of a string array with the values of all action units to the Alfred application.



Figure 4.3: The virtual character Alfred is designed utilizing FACS to compose facial expressions.

4.2.3 Design and Implementation of the Interface

We connected this facial animation system to three controllers: (1) a slider-based GUI, which is the current standard interface for such a task; (2) a gamepad, which allows parallel control and is widely used in computer games; (3) and a data glove, which enables continuous control while editing five facial parameters in parallel. We first conducted a survey with professional graphics designers to find out how the latter two new input devices would be received in a commercial context. A second study with 17 subjects was conducted to analyze the performance and quality of these two new input devices.

We identified a number of serious disadvantages regarding wide spread slider-based user interfaces for facial expression generation. Hardware controllers represent a promising alternative which has not yet been explored in depth. We analyzed the capabilities of such input devices and then defined an intuitive mapping between the input devices and the facial expression control.

Use of Novel Input Devices

Slider-based GUIs limit the composition of facial expressions to sequential control and thus lack transparency. Since users can just edit a single facial unit at a time, it is hard for them to imagine for what the final result of the composing might look like. Novel input devices, such as gamepads or data gloves, allow users to modify several facial units at once.

Gamepad The first type of hardware controller we studied in our work was the gamepad. Gamepads are today's standard controller for gaming consoles like the Xbox 360 or the Playstation 3. They have the major advantage of being widely available, cheap, and many users are familiar with them. Gamepads were originally designed for long hours of computer gaming and thus their design takes into account many ergonomic aspects. We focused on the Xbox 360 game controller (see Figure 4.4), which can also be easily connected to any Windows compatible PC. Since most of today's gamepads are constructed in a similar manner, our analysis and results can be easily transferred to other gamepads.



Figure 4.4: The Xbox 360 controller with two analog sticks, two analog shoulder buttons, one digital stick and several buttons.

To control facial expression, it is important that a controller returns a continuous

data stream. In this way, the intensities of the action units or morph targets can be controlled in real-time. The XBox 360 gamepad provides a variety of analog and digital controls: two analog sticks, one four-way digital cross, six buttons on the top of the gamepad and four buttons, two of which are analog, on the front of it (see Figure 4.4). Each of them can be controlled independently and in parallel by moving a finger or thumb. The analog buttons provide a one dimensional signal similar to sliders and the analog sticks provide a two dimensional signal. Two basic approaches should be considered to interpret this two-dimensional signal and to transfer it into the one-dimensional “action unit”-space:

1. The signals from the analog sticks consist of two dimensions: an x- and a y-dimension. Each dimension of an analog stick can be mapped to one parameter of an action unit. In this way, two action units can be controlled simultaneously. But, contrary to sliders or analog buttons, analog sticks provide positive and negative values. This allowed us to map negative values into a positive space and thus control four different parameters with one analog stick (i.e. moving the analog stick forward to control one action unit and moving it backward to control a second action unit – the same when moving the stick sideways).
2. Since analog sticks can be moved circularly, signals can also be interpreted as polar coordinates. In that way, the angular coordinates can be used to select an action unit and the radial coordinate can be used as its weight.

The first approach controls two action units at once, since the horizontal and vertical activations are independent. With the second approach, only one action unit can be controlled at once, as every angular activation selects an action unit. In addition to the analog controls, the XBox 360 gamepad has a couple of digital buttons and a directional pad, which can be used for further control functions (e.g. switching the current setting of action unit mapping).

Data Glove Data gloves (see Figure 4.5) measure the bends of the fingers and, often too, the orientation and the position of the hand wearing the data glove. While the position of the hand can be very useful for performing a selection task (e.g. selecting the setting for a certain region of the face), the posture of the hand can be used for expression control. The human hand consists of five fingers, which can be bent relatively independently. Since a finger bend is a one-dimensional signal, it is an ideal candidate to replace slider-based interfaces.



Figure 4.5: The P5 data glove with five analog controllers and three buttons.

The “P5 Glove” was originally developed for gamers, and its low cost makes it ideal for developing prototypical applications. The “P5 Glove” provides the following data:

- absolute position (x,y,z) , relative position (x,y,z) , and rotation (yaw, pitch, roll)
- finger bend
- three additional digital buttons

Mapping Models

The question arose as to how the single signals of a controller could be projected onto the FACS model. Both the gamepad and the data glove offer just a limited number of controls which do not suffice to cover our 23 action units. In this section, we present three different mapping models to solve this problem.

Direct Mapping The basic idea of direct mapping is to transfer the structure and layout of the human face onto the hardware controller. To this end, the face is



Figure 4.6: Settings for the gamepad to control the action units for the upper face (left), lower face without inner lips (middle), and the inner lips (right).

decomposed into logical groups (e.g. eyes, nose, mouth etc.). This methodology is already defined by Ekman and Friesen (1978), which describe logical groups for facial regions. They decomposed the face into an upper part with 7 action units and a lower part with 16 action units.

Since there are more action units than may be controlled at a time, it was necessary to assign multiple settings to one control. Suitable controls are those that are not required for manipulating facial action units. Here, principles of ergonomics should be applied. Controls which are easy to operate should be reserved for the more important facial action units. The importance of an action unit is defined by its frequency of occurrence and its influence on the facial expression.

Mapping for Gamepad Based on the considerations above, we defined the following settings for the gamepad. The upper face with 7 action units could be directly mapped to a gamepad setting with two action units mapped to the two front buttons, two action units mapped directly to the right analog stick and three action units mapped to the left analog stick using circular mapping. The lower face had to be split into two settings, since 16 action units represented too many regulators for the gamepad to be able to deal with them all at once. The second setting controlled parts of the lower face, excluding the action units for the inner lips. We again used the analog front buttons to directly map two action units. The left analog stick was used with circular mapping to control the lip corners and the right stick was mapped to the raising and lowering of the chin. The third setting controlled the inner lips. The front buttons were used to control two action units, the left stick to control one action unit and the right stick to control two action units. Figure 4.6 illustrates the gamepad settings for the upper face, the lower face, and the inner lips. The four-way digital cross is used to switch between the three settings.

Mapping for Data Glove The data glove provides five analog controls, one for each finger. That means that only five action units can be controlled simultaneously. The 23 action units could be mapped to five different settings for the data glove. To keep Ekman and Friesen (1978) distinction between upper and lower face, we opted to use six logical settings for the data glove: brows (3 AUs), lids (3 AUs), cheek and nose (3 AUs), corners of the mouth (5 AUs), chin and inner lips (4 AUs), and lips (3 AUs). More important action units were mapped to more prominent fingers. The user can select one of the six settings by moving the data glove horizontally.

Context-Sensitive Mapping One disadvantage of direct mapping is the necessity to assign multiple functions to a single control. Permanently switching between different settings increases the complexity of the interface and thus the time required to generate a facial expression. To avoid this problem, we investigated whether it would be possible to have the user only manipulate action units that are relevant in a specific context. An example of such a context would be an emotion the user wishes to express. In such a situation, it might be helpful to provide the user with just the action units that are necessary to adjust the corresponding emotional expression as desired. We conducted an analysis of the facial expressions stored in the FER database in order to find out which action units were mostly involved in a particular facial expression considering their frequency of occurrence as well as the variance of occurrence. In addition, we performed a correlation analysis in order to identify action units occurring together.

The FER database contains variations of Ekman's basic emotions: joy, anger, fear, sadness, disgust, and surprise. We use these six emotional expressions as the context to be modified. The action units to define these basic expressions were selected by collating the listed action units from the FER database with the results presented by Ekman and Friesen (1978) and Kätsyri (2006). The overlapping action units were used to define a basic emotion. The action units that influence a basic emotion (e.g. surprise → puzzled) were selected by calculating the mean and the variance. Action units with a high mean value were considered important for varying the basic emotion and were automatically mapped to an analog control on the gamepad. Action units with a medium mean value but a high variance were considered important for a broader range of different facial expressions and thus were mapped to an analog stick using the circular approach. Action units with a low mean value and variance were checked manually to see whether they played an important role in influence the facial expression of a basic emotion and, where applicable, were omitted.

Using the gamepad to modify a basic emotion, the user could select the desired

context by pressing the four-way digital cross. The respective basic facial expression was presented and the user could manipulate this expression using a setting which was adjusted to this particular context. For instance, if a user selected the context “joy”, the setting contained, among other modifiers, the action units “cheek raiser” and “lip corner puller”, whereas the action unit “lip corner depressor” was not allocated to a control in this setting, as it was not needed.

Mapping Based on Basic Emotion Categories Ekman and Friesen (1975) found that a large portion of emotional expressions could be generated by blending the six basic emotions. McCloud (2006) seized on this idea and illustrated how comic characters can express emotions by blending two or more of the six basic emotions.

We implemented two mappings for blending the basic emotions with the gamepad. One uses the six independent analog controls and maps the intensity of all action units for one basic emotion to one control of the gamepad. The user simultaneously controls all six basic emotions with four fingers, two emotions for each analog stick and one emotion for each analog front button. This approach might be challenging for the user, as all six emotions can be blended at once. McCloud (2006) mostly blends two basic emotions, thus it could be sufficient to limit the controls to blending two to four basic emotions at once. The second mapping uses the two analog sticks of the gamepad based on the circular approach. This allows the user to blend two to four basic emotions using two controls simultaneously.

Although the blending on face level produces a variety of different emotional facial expressions, blending on face regions would not only increase the variations, but also improve the quality of such blended expressions. Especially, if emotions overlap (e.g. you feel sad but want to show joy), the way on how treat the blending on the different facial regions is challenging (Ekman and Friesen, 1975) (Ochs et al., 2005). In this chapter, we did not consider the blending at the level of facial regions, as we wanted to keep the blending of facial expressions as simple as possible in this first approach.

Anatomical Constraint Model

One problem with morph targets are the interferences that might occur when simultaneously activating several morph targets. (Lewis et al., 2005) describes this phenomenon and offers a solution to avoid such interferences. Since the FACS model was originally defined to analyze, and not generate facial expressions, it is possible to simultaneously activate certain action units, which anatomically

speaking would be impossible, and the result of this is an unnatural facial expression (see Figure 4.7).

Our constraint model, to prevent such unnatural facial expressions, is based on facial regions. Action units that are anatomically impossible within one facial region are reduced to a realistic value.

$$Totalforce_{region} = \sum_i AU_i * weight_{i,region}$$

When the total force of a region exceeds 1.5, all action units within this region are reduced by the factor $Totalforce/1.5$:

$$AU_{i,reduced} = \frac{1.5 * AU_i}{Totalforce_{region}}$$



Figure 4.7: Facial expression without (left) and with (right) constraint model.

The face is divided into four regions, which are not dependent on each other: eye brows, eye lids, inner lips, and corners of the mouth. The single weights for each action unit were derived from video clips in the FER database and manually adjusted. Although the constraint model is in principle independent from the facial model as it is based on the activation of the AUs, it might depend on how the single morph targets for the AUs are designed.

4.2.4 Studies with Professional Graphics Designers

To find out how the new interfaces would be received in a commercial context, we recruited two professional graphics designers from the computer game developer “Chimera Entertainment”. The study served to clarify a number of questions that came up during the design of the interface. In particular, we hoped to get useful hints regarding the assignment of functions to hardware controls. The introduction to the system followed the Coaching Method in order to obtain additional information on the users’ behavior during the learning phase. During the actual test, the users performed different tasks following the “Thinking Aloud” method.

Our users appreciated the direct gamepad control interface. In particular, they found that this interface had a clearer structure and layout than the slider-based interface. Regarding the context-sensitive gamepad control interface, a number of concerns were uttered. Firstly, the continuous switch between different settings required the users to re-orient themselves again and again and made it difficult for them to get familiar with the interface. Secondly, the users had the feeling that they had less control over the system as a whole since it was not always obvious to them which action units were to be manipulated.

The basic emotion composition approach was positively received. This approach was described as intuitive and easy to use. In particular, the participants appreciated the fact that this approach could speed up the production process. The participants, however, had some doubts as to whether it would be possible to adjust the settings in such a way that all desired emotional expressions could be generated. Nevertheless, they regarded this approach as a solution to come up with fast and creative pre-settings. In particular, the potential of the composition approach in combination with direct mapping was emphasized. A designer could, for instance, first create a rough pre-model and then refine it using the direct mapping approach.

The data glove profited in particular from the novelty effect. As graphics designers, our users were familiar with gamepads, but not with data gloves, which are less common in the game industry. Yet, they found that the data glove was not accurate enough. Furthermore, it was perceived as physically tiring after some time.

The users emphasized the importance of comprehensive functions for storing and changing authored expressions for day-to-day production. Moreover, they mentioned the noisy signal from the data glove regarding the tracking of the hand in space. They found it quite difficult to select a setting, which was mapped to the horizontal movement of the glove. To improve the selection process a noise

reduction filter was applied to the signal from the glove.

At first, there were two versions of the gamepad visualization. One was with text that labeled the controller with the controllable action unit (see Figure 4.6) and the other was with small icons of the controllable action unit. The users preferred the one with text and therefore the icons were omitted.

4.2.5 Evaluation of the Hardware Controllers

To compare the novel hardware devices against the traditional sliders, we conducted a formal user study. We were particularly interested in finding out (1) how users would get along with the novel input devices in comparison to the sliders, (2) whether they would enjoy using them and (3) how they assessed their technical features. Besides subjective user ratings, we also aimed at objective performance measurements. In particular, we assessed the quality of the users' creations and recorded how much time it took them to accomplish a task.

Based on the preliminary user study, we expected the novel input devices to be positively received. We assumed that both the data glove and the gamepad would contribute to an enjoyable interaction experience. In particular, we believed that the gamepad would successfully compete against the sliders thanks to better usability and performance. Due to the feedback we got from the professionals, we did, however, expect some usability and performance issues with the data glove.

The formal study was structured as follows: Each input device was tested by each participant in random order. After a short training phase, we presented the participants with concrete modeling tasks that they had to accomplish using a particular input device and measured task completion in terms of quality and time. Before the participants were given the tasks for the next input device to be tested, they were asked for their subjective assessment of the input device they had just used.

Users and Experimental Method

We recruited 17 subjects aged 20 to 40 (13 males and 4 females). Most of the subjects (76 %) were students. To assess participants' prior experience, we used a 5-point rating scale: "none", "little", "medium", "high" and "extremely high". Most participants were familiar with the use of gamepads (mean value: 2.82), but had in general little experience with 3D modeling (mean value: 1.88), facial expression animation (mean value: 1.47) or emotional models (mean value: 1.88). Due to the large number of users required for a statistical analysis, it was not possible to rely on professional graphics designers for the formal user study. The

different user groups, however, gave us the opportunity to investigate whether the comments of the professionals could be confirmed by non-professionals.

At the beginning of the experiment, the participants had to input the required demographic data. After that, they tested each input device in random order to avoid any bias due to habituation effects. To compare the single input devices, we decided to rely on the direct mapping approach. First of all, the interviews with the professionals had revealed a preference for the direct model approach. Secondly, it would have been more difficult to identify the factors responsible for a particular effect using the other two mapping models since they heavily differed for the single input devices. For each input device, the participants had to go through the following phases:

- *Training Phase*

The single participants were given an individual introduction to the input device to be used next while they were holding it in their hands. After that, they got one minute to test the input device themselves.

- *Modeling Phase*

The participants were asked to create three facial expressions based on photos of an actor. The photos were taken from the FER database which provides a list of the relevant action units for each facial expression. To test the single input devices with different photos, we collected nine photos that were distinguished by different levels of complexity:

- Facial expressions with *complex eye area* and *simple mouth area*
- Facial expressions with *simple eye area* and *complex mouth area*
- Facial expressions with *complex eye area* and *complex mouth area*

The complexity was defined by the number of action units that had to be manipulated in order to create a particular facial expression. The facial expressions were randomly assigned to the three input devices. Using a particular input device, the participants had to generate one facial expression per category. They were allowed to spend as much time as they wished on a specific modeling task. While they were interacting with the system, the time was logged. After each task, participants were asked to indicate how satisfied they were with their result using a questionnaire with a 5-point scale attitude statement (disagree, somewhat disagree, neutral, somewhat agree, agree) for each task.

- *Questionnaire*

After accomplishing all three tasks with a particular controller, the participant was asked to fill in a post-task questionnaire (see Appendix C.1). The post-task questionnaire used eight attitude statements with a 5-point scale to evaluate how the participants perceived the interaction with the system when using a particular device. The question referred to the usability of the device (four questions: U1, U2, U3, U4), the user's subjective perception of the interaction experience (two questions: E1, E2) and the technical features of the device (two questions: T1, T2).

Finally, the participants were asked which controller they would choose if they had to repeat the test again with all nine photos.

Results of the Experiment

To evaluate the gamepad and the data glove, we compared them with the sliders as a reference interface. In particular, we applied two-tailed t -tests to each of the two novel input devices and the sliders.

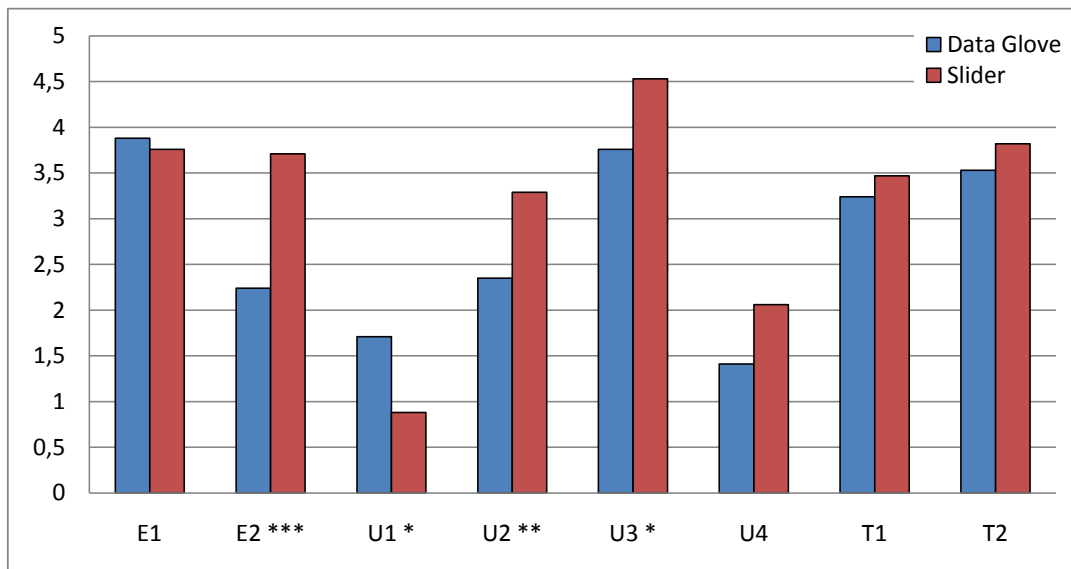


Figure 4.8: Subjective user ratings for the Data Glove compared to the Slider based approach (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

First we analyzed how the participants had assessed the results of their own work. Overall, the participants were most satisfied with the facial expressions they created using the sliders with a mean value of 3.84, followed by the facial expressions they created using the gamepad with a mean value of 3.63. The data glove scored

worst with a mean value of 3.30. Significant differences were only found for the data glove and the sliders ($p = 0.018$).

Having participants assess their own results is, however, a subjective quality measurement. Since it is unclear which criteria the participants used and what factors influenced their ratings, such data should be interpreted with caution. We therefore decided to complement the participants' subjective ratings by objective quality measurements. In particular, we computed to what extent a facial expression created by the subjects deviated from a reference facial expression. To this end, we created for each of the nine tasks a standard expression based on the action units that were listed for the corresponding photo in the FER database. We then calculated the deviation of the facial expressions created by the participants from the corresponding standard reference expression using the following formula

$$Deviation = \sum_i |AU_{ref,i} - AU_{user,i}|$$

where AU is a floating value between 0 and 1 and i is the index of all action units. Using this formula, we obtained the following mean values for the deviation of user-generated facial expressions from the corresponding standard expressions: 4.26 for the gamepad, 4.53 for the sliders and 4.94 for the data glove. Neither the value for the gamepad nor for the data glove was significantly different from the value for the sliders.

When analyzing the time the participants spent on the creation of facial expressions, we found that they needed significantly less time with the gamepad (148.1 s) than with the sliders (168.3 s), while they needed significantly more time with the data glove (263.3 s). The time advantage for the gamepad was of about 12 % averaged over all values. For six out of seventeen users, the time advantage for the gamepad was even above 30 %, while just one user got a time advantage of above 30 % for the sliders. However, only the difference between the values for the data glove and the values for the sliders were significant ($p < 0.0005$), while the difference between the values for the gamepad and the values for the sliders were just tendentially significant ($p = 0.056$).

In addition to evaluating the performance of the single devices in terms of quality and time, we were interested in the participants' subjective impression. Overall, the gamepad achieved the best mean scores for most attitude statements. Figure 4.8 shows the results of two-tailed t -tests for the data glove compared with the sliders and Figure 4.9 shows the results for the gamepad and the sliders. Results that were statistically significant are marked by stars. The findings of the experiment may be summarized as follows:

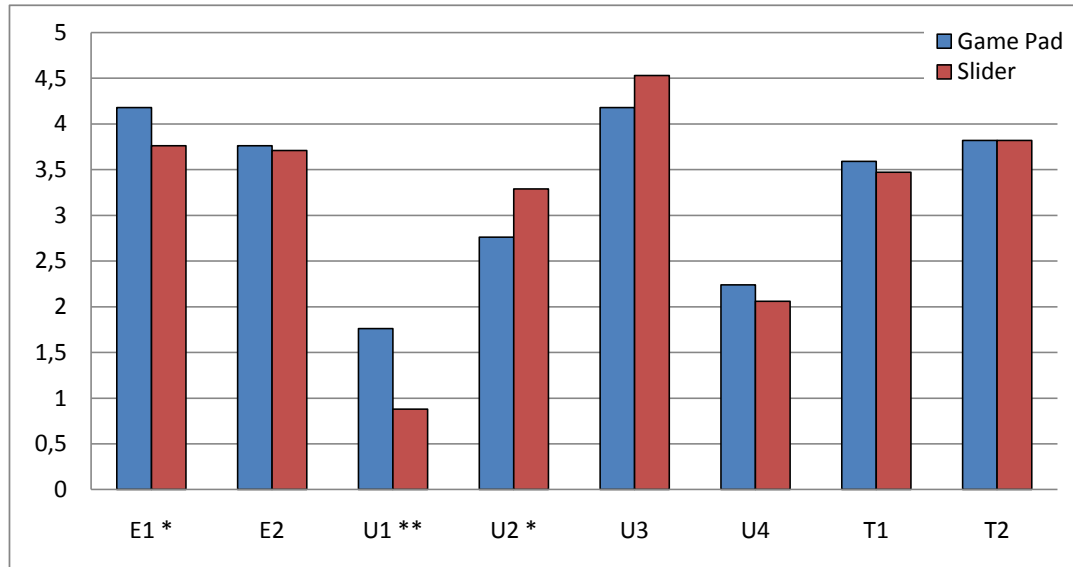


Figure 4.9: Subjective user ratings for the Gamepad compared to the Slider based approach (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

- *Interaction Experience:* The participants found it more enjoyable to use the gamepad and the data glove than to use the sliders (E1). However, they found the data glove physically more tiring than the sliders (E2).
- *Usability:* A major advantage of the novel input devices in comparison to the traditional sliders is that they allow the users to keep their eyes on their work. In the experiment, the participants had the feeling that they had to shift their gaze more often between the input devices and the characters when using the sliders than when using the gamepad or the data glove (U1). The sliders, however, scored best regarding the predictability (U2) and the plausibility (U3) of the devices' behavior. Compared to the sliders and the data glove, the gamepad enabled better tuning. The participants found it less difficult to adjust parameters with the gamepad than with the other two devices (U4).
- *Technical Features:* The participants had the impression that the gamepad offered them more options to adjust parameters than the other two devices (T1) and were more satisfied with the accuracy of the gamepad than with the accuracy of the data glove (T2).

Discussion

Overall, the use of a gamepad for facial expression generation can be regarded as promising. It reduced the production time without causing a loss of quality. This

result is all the more remarkable as the gamepad hardware is obviously not adjusted to the specific requirements of facial animation design. Thus it came as no surprise that a large proportion of our users (49 %) expressed a preference for the gamepad. The sliders only scored better regarding predictability and plausibility, which could be explained by the fact that the sliders were labeled with the action units the user could manipulate.

The bad score of the data glove deserves further discussion. Even though data gloves were recommended as an input device for facial animation already by *Thalmann (1993)* and furthermore used in production by Quantic Dreams, they obtained a significantly lower rating on almost all attitude statements. Furthermore, it took our users significantly longer to come up with a result than with any of the other devices, and the quality of the result was significantly lower. The only advantage found over traditional input devices was that data gloves allow the users to direct their gaze fully onto their work and do not require them to permanently shift their gaze between the interface device and the graphical display. The low wearing comfort and insufficient accuracy of the very low-priced hardware may explain the poor ratings. Furthermore, moving five fingers in parallel might have been too difficult for unexperienced users. Finally, pressing buttons with the left hand was most likely too complicated and caused an interruption of the work flow. Nevertheless, the data glove should not be discarded as a completely inoperative input device. After all, 24 % of the users preferred it as a controller – nearly as many as those who chose the definitely superior sliders. One of the users indicated that the data glove offered him the maximum amount of parallel control over the facial action units.

4.3 Conclusion

This chapter introduced the Horde3D GameEngine and presented the required components for enabling a face-to-face communication between a user and a virtual character's head. The Horde3D GameEngine contains components for controlling the facial expressions of a virtual character, to handle the lip synchronization while speaking, an inverse kinematics component for controlling the head orientation and gaze of a virtual character and 3D object detection, which allows to detect where the user is currently looking at in the 3D scene.

Further, we investigated three different interfaces to a FACS-based animation system. Based on ergonomic principles, we defined three mapping strategies to assign facial actions to controls and showed how they could be applied to gamepads and data gloves. The appropriateness of the mapping strategies was investigated

by conducting interviews with professional graphics designers. Based on these studies, we tested the most promising mapping strategies for the gamepad and the data glove in an experiment with non-experienced users. The users had to accomplish various tasks which were evaluated based on time and quality. The users were not only satisfied with the facial expressions they created. In addition, there was a high congruence between the users' creations and the corresponding standard reference expressions. A comparison of the novel hardware devices with the conventional sliders revealed that the gamepad scored best on most dimensions. It helped reduce production time without loss of quality.

Chapter 5

Study II: Perception of Facial Display, Head and Eye Orientation

In this chapter we focus on the perception of facial expressions, head and eye orientations of virtual characters in combination with the expression of social dominance. In particular, we are interested in the interaction of different non-verbal cues. We present a study which systematically varies gaze and head tilts for five basic emotions and a neutral state using our own graphics and animation engine. The resulting images are then presented to a large number of subjects via a web-based interface who are asked to attribute dominance values to the character shown in the images. First, we analyze how dominance ratings are influenced by the conveyed emotional facial expression. Further, we investigate how gaze direction and head pose influence dominance perception depending on the displayed emotional state (Bee et al., 2009c).

The second part of this chapter focuses on the perception of facial expression, head and eye orientations of virtual characters in combination with the expression of the three personality traits extraversion, agreeableness and emotional stability. We present a study which systematically varies head orientations and gaze to measure the personality perception. We conducted a web-based study with 133 persons to rate 54 images of a virtual character with varying head and gaze directions. First, we analyzed how the different head directions influence the three measured personalities traits: extraversion, agreeableness, and emotional stability. We chose the Big-Five Factor model (Goldberg, 1992) to measure the perception of personality as it is well established in research work on personality perception. Further, we investigated how head pose and gaze direction influence personality perception depending on the way they are combined (Arellano et al., 2011).

5.1 Experiment I: Dominance Perception

In order to come across as believable, virtual agents need to portray social behaviors in a convincing manner. Among other things, social behaviors are reflected by the way a character communicates social dominance. Prior psychological experiments indicate that the following facial cues are used to express social dominance: facial expressions, gaze and head tilts.

The first experiment will investigate how to orchestrate facial expressions, gaze and head tilts when a virtual agent is expected to express social dominance. We use a virtual character whose facial expression is controllable through FACS (Facial Action Coding System) (Ekman and Friesen, 1975). To analyze social dominance, we apply methods from Mehrabian and Russell (1974), who defined the PAD (*Pleasure, Arousal and Dominance*) model for emotions by using bipolar pairs of words to judge humans' affective attitudes towards different situations (see also (Mehrabian, 1995)). Based on the PAD model, Russell and Mehrabian created a dictionary with 151 emotional terms mapped onto pleasure, arousal and dominance (Russell and Mehrabian, 1977).

In the following, we will present an experiment we conducted in order to investigate how facial expressions, gaze and head tilts of a virtual character influence the perception of dominance. On the basis of related studies, we expected the following outcomes:

- Dominance perception of a virtual human is influenced by the facial display of an emotion. In particular, facial expressions that convey joy, anger and disgust are perceived as more dominant than neutral, fearful or sad facial expressions.
- The perception of dominance depends on the direction of the gaze and head orientation. In particular, people who avert their gaze and look down appear to be more submissive than people using direct gaze and looking up.
- There are interactions between emotional facial displays, head and gaze direction. In particular, we assume that direct gaze in combination with upward head orientation will increase dominance ratings for joy, anger and disgust while averted gaze in combination with downward head orientation will decrease them. Vice versa, we expect that averted gaze in combination with downward head orientation will decrease the dominance values for fear and sadness while direct gaze in combination with upward head orientation will increase them.

5.1.1 Affective and Attentive Virtual Character

For our study, we used a fully controllable virtual head named Alfred (see Figure 5.1 and also Section 4.2.2), our butler-like character. Alfred uses action units to synthesize a huge amount of different facial expressions. The action units are designed using morph targets and thus give a designer the full power in defining the facial expression outlook. The system includes a tool to control the single action units, which enables us to store the result in an XML file for later usage in our agent system (Bee et al., 2009b) (see also Section 4.2).

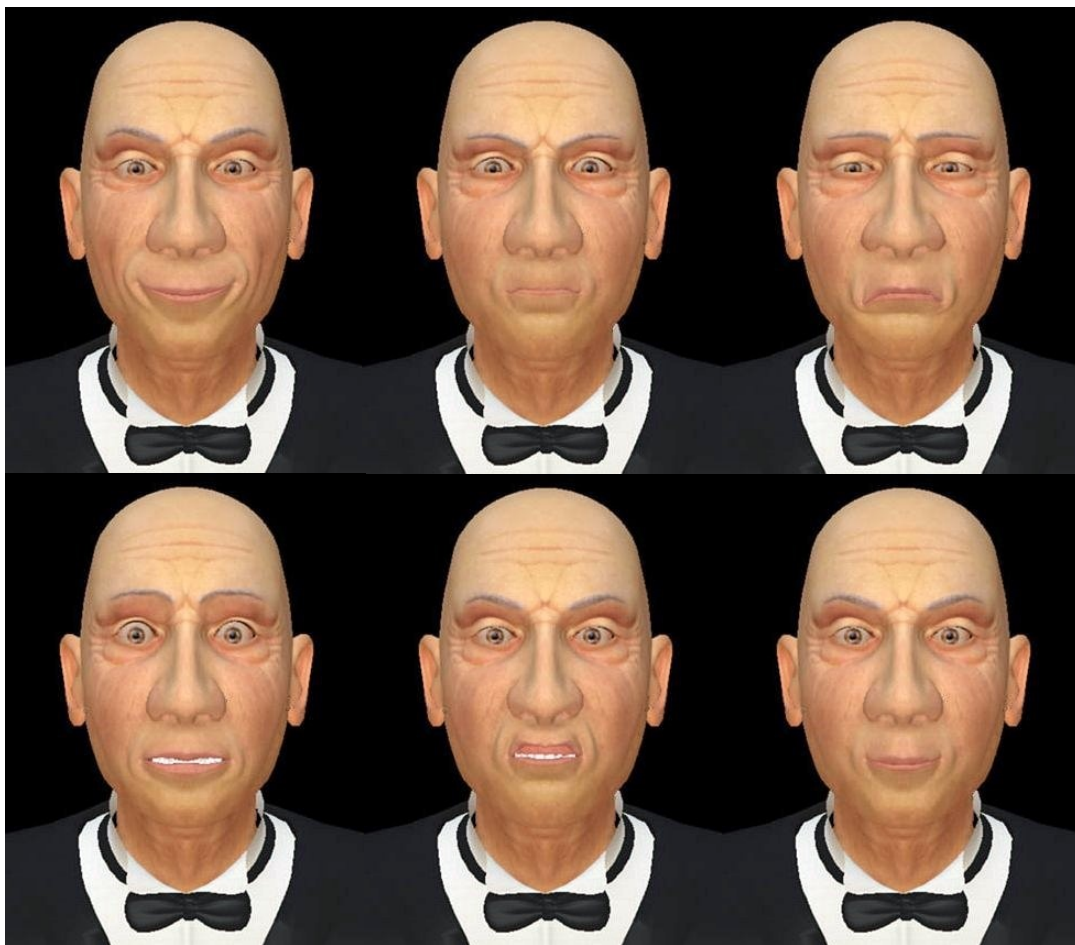


Figure 5.1: The six facial expressions of Alfred used in the study. (Upper row from left to right: joy, anger and sadness; lower row: fear, disgust and neutral)

The Horde3D GameEngine developed by (Augsburg University, 2007) provides an inverse kinematics component (see Section 4.1.3) with which the head and eye direction can be controlled. The head and eye direction can be set independently by the IK component (e.g. the head can look to the right while the eyes are directed towards the center).

5.1.2 Experimental Study

The parameters for the affective display of the virtual character were picked from the (Facial Expression Repertoire, 2008). The FER provides samples of over 150 facial expressions that may be mapped onto the action units of FACS. The database does not only indicate which action units have to be activated for certain facial expressions, it also provides a rich dataset of videos which show how the action units ought to be designed. We chose the following representatives: Joy (action units 6 and 12), Anger (action units 4, 5, 7, 17 and 23), Sadness (action unit 1, 4, 7, 11, 15 and 17), Fear (action units 1, 2, 4, 5, 20 and 25) and Disgust (action units 10 and 17).

We varied the attentiveness of the virtual character by modifying head orientation and gaze direction. The maximum angle for the gaze was defined by the limits of the pupils' visibility and also applied for the head orientation. This led to nine gaze orientations for the head and eyes (up (N), center (C) and down (S) varied by $\pm 8.0^\circ$ and left (W), center (C) and right (E) varied by $\pm 8.5^\circ$).

The possible combinations of 6 emotions \times 9 head directions \times 9 eye directions result into 486 different facial displays. To reduce them to a reasonable amount, we assumed that it does not matter whether the virtual character gazes to the left or to the right and randomly distributed the horizontal gaze to one of these sides. To further limit the amount of facial expressions, we removed the facial expressions where the eye is gazing into the opposite direction of the head orientation (i.e. the head is directed to the left and the eyes are directed to the right). These two assumptions limit the number of possible facial displays to 194. We automatically generated pictures from these settings with a script which controls Alfred and automatically saves the expression as a screenshot.

To obtain the dominance values for the affective and attentive facial expressions, we followed the instructions from Mehrabian and Russell (1974). They provide 18 pairs of words, 6 for each of the dimensions (i.e. pleasure, arousal and dominance), which need to be rated on a 9-point scale. As we were mainly interested in the dominance factor of the facial displays, we limited these pairs of words to the six that are necessary for obtaining the dominance factor (i.e. *Controlling - Controlled*, *Influential - Influenced*, *In control - Cared-for*, *Important - Awed*, *Dominant - Submissive* and *Autonomous - Guided*; see Appendix C.2).

Overall 69 (40 female and 29 male) participants judged in total 862 pictures. The mean age was 28.5 and the participants came from all walks of life. Each of the seven emotions was judged about 123 times and each of the 194 pictures was judged 4.4 times on average, whereby every picture was judged at least 4 times, but

maximally 5 times. On average, the subjects rated their experience in 3D modeling with 0.49, their experience in animating facial expressions with 0.23 and their background in emotion research with 0.36 on a scale between 0 and 4 (with 0 representing no experience at all). That is our subjects had no experience in any related field.

5.1.3 Results

The analysis of dominance ratings for the different combinations of gaze, head pose and emotional displays was based on the one-way analysis of variance (ANOVA) across the different groups and the Tukey-HSD for the post-hoc two-sided pairwise comparisons. *t*-Tests were applied two-sided, as we did not predefine, in which direction the means differed.

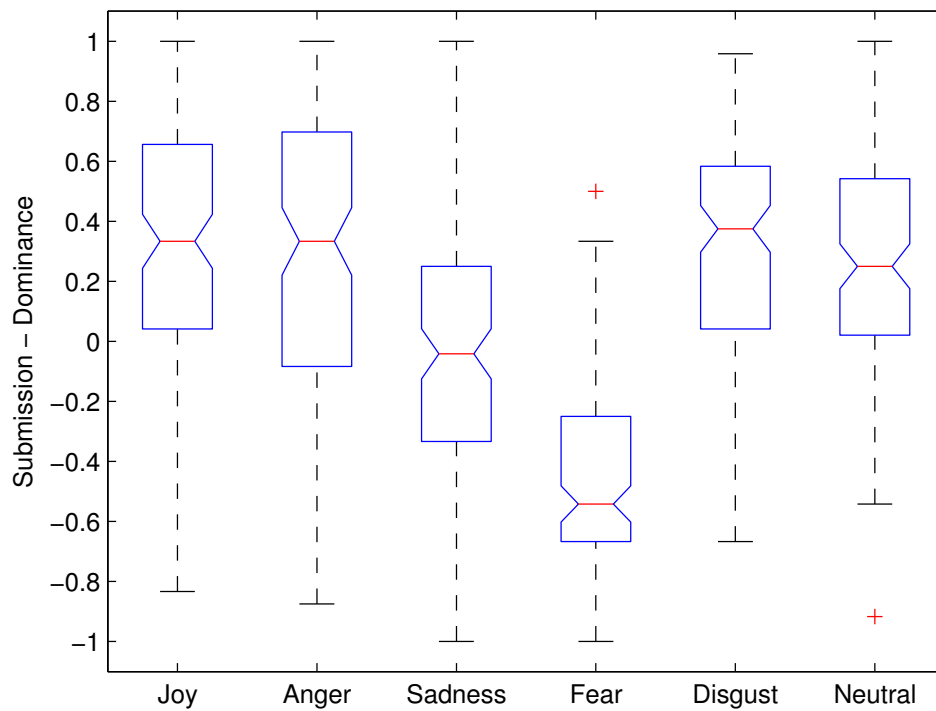


Figure 5.2: Dominance values with all eye and head orientation variations

Influence of Emotional Displays, Gaze and Head Tilts on Dominance Perception

Influence of Emotional Displays on Dominance Perception. A comparison of all six facial expressions including all variants of the gaze and head orientations

	Russell & Mehrabian	Our Experiment
Joy	0.35	0.32
Anger	0.25	0.29
Sadness	-0.33	-0.04
Fear	-0.43	-0.46
Disgust	0.11	0.31

Table 5.1: Comparison between Russell and Mehrabian’s and our dominance values

with ANOVA revealed significant differences in dominance perception among the emotion groups ($F(5, 705) = 74.6, p = 0, \eta^2 = 11.9$). The pairwise Tukey-HSD post-hoc analysis revealed significant differences between all groups, except *Joy – Anger*, *Joy – Disgust*, *Joy – Neutral*, *Anger – Disgust*, *Anger – Neutral* and *Disgust – Neutral* (see Figure 5.2 and Table 5.2).

When comparing our results to the dominance values for the emotions Joy, Anger, Sadness, Fear and Disgust by Russell and Mehrabian (1977), we noticed only small differences for Joy, Anger and Fear (see Table 5.1).

Furthermore, most of our results are in line with a study previously conducted by Knutson (1996). Joyful, angry or disgusted facial displays were rated more dominant than fearful and sad facial displays. As Knutson, we got the highest dominance value for joy and the lowest dominance value for fear. However, unlike Knutson, we observed that neutral facial expressions were perceived nearly as dominant as joyful or angry facial expressions.

Influence of Gaze on Dominance Perception In the previous paragraph, the influence of gaze and head orientation on dominance ratings was not separately analyzed. In the following, we will investigate how the gaze influences the perception of dominance whereby we will compare the effect of directed with the effect of averted gaze.

Dominance of direct gaze The one-way ANOVA revealed significantly different dominance ratings within all six groups of affective displays ($F(5, 19) = 7.96, p = 0, \eta^2 = 1.04$) when the eyes and the head were directed at the user. A post-hoc analysis showed significant differences between several emotions. The Tukey-HSD analysis revealed that *Joy* significantly differed from *Anger* ($p < 0.01$), *Fear* ($p < 0.001$) and *Sadness* ($p < 0.05$). And further, it revealed significant differences

between *Fear* and *Disgust* ($p < 0.01$) and between *Fear* and *Neutral* ($p < 0.05$) (see Figure 5.3).

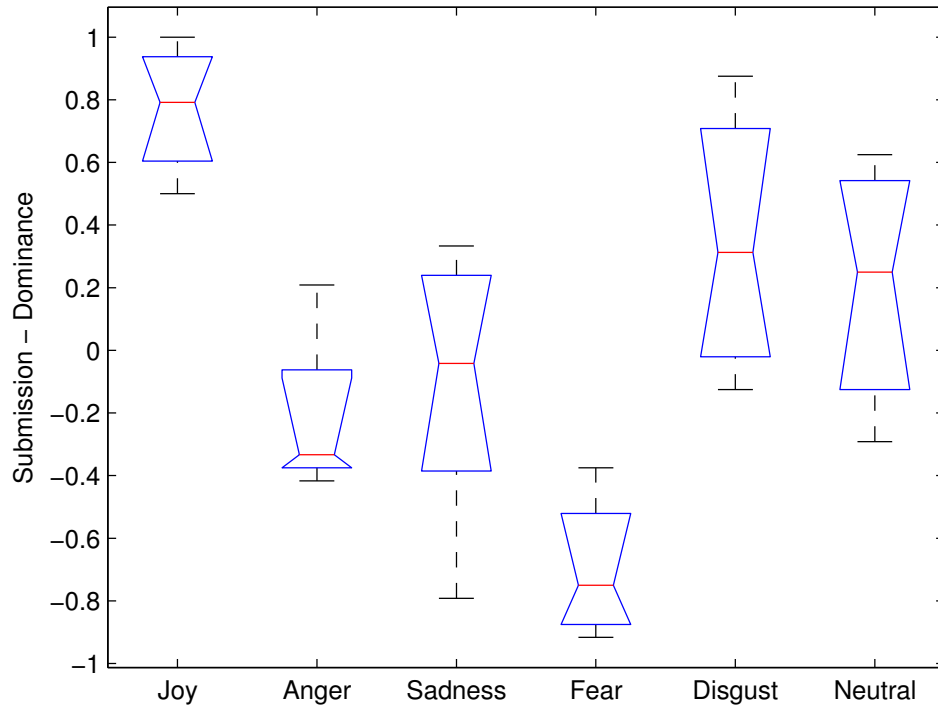


Figure 5.3: Dominance values with eyes and head directed at the user

Dominance of Averted Gaze The one-way ANOVA for averted eye/head gaze revealed significantly different dominance values within the groups of all six affective displays independent from the direction of the averted gaze ($F(5, 680) = 71, p = 0, \eta^2 = 11.3$). A post-hoc analysis shows significant differences between several groups of emotions. Table 5.2 shows the results for the Tukey-HSD post-hoc analysis between the different emotion groups, where, for example, sadness or fear significantly differs from all the other emotion groups.

Direct vs. Averted Gaze A comparison of averted and direct eye/head gaze revealed significant differences for Joy and Anger. The dominance value of Joy dropped from 0.80 for directed gaze to 0.33 for averted gaze. A two-tailed t -test revealed that the dominance value for averted gaze was significantly lower ($t = -4.1, df = 4.1, p < 0.01$) than the dominance value for direct gaze. In contrast the dominance value for Anger rose from -0.33 in the case of direct gaze to 0.38 in the case of averted gaze. A two-tailed t -test revealed that the dominance value

	Joy	Anger	Sadness	Fear	Disgust	Neutral
Joy	–	n.s.	***	***	n.s.	n.s.
Anger	n.s.	–	***	***	n.s.	n.s.
Sadness	***	***	–	***	***	***
Fear	***	***	***	–	***	***
Disgust	n.s.	n.s.	***	***	–	n.s.
Neutral	n.s.	n.s.	***	***	n.s.	–

Table 5.2: Post-hoc comparisons for averted gaze and head orientation between different emotions (*** $p = 0$, n.s. = not significant)

of averted gaze was significantly higher ($t = 3.6$, $df = 4.1$, $p < 0.05$) than that of direct gaze. Fear ($p = 0.1$), Sadness ($p = 0.57$), Disgust ($p = 0.9$) and the Neutral ($p = 0.86$) affective display were independent from the virtual character’s current eye/head gaze (see Table 5.3). They did not show any significant differences for averted versus directed eye/head gaze.

	Direct	Averted
Joy	0.80	0.33
Anger	–0.33	0.38
Sadness	–0.04	–0.04
Fear	–0.75	–0.52
Disgust	0.31	0.38
Neutral	0.25	0.25

Table 5.3: Median values for dominance over all emotions with directed eye and head direction compared with averted eye and head direction

Influence of Gaze in Combination with Facial Displays on Dominance Perception

As head orientation dominates gaze, we keep the head oriented towards the user and limit our analysis here to different gaze directions.

In Section 5.1.3, significant differences between averted and directed gaze were observed for Anger, Fear and Joy. When we varied the direction of gaze with the head oriented directly towards the user, the one-way ANOVA revealed significant differences in dominance perception for Anger ($F(5, 20) = 4.2$, $p < 0.001$, $\eta^2 = 0.49$). The post-hoc Tukey-HSD revealed only significant differences between the

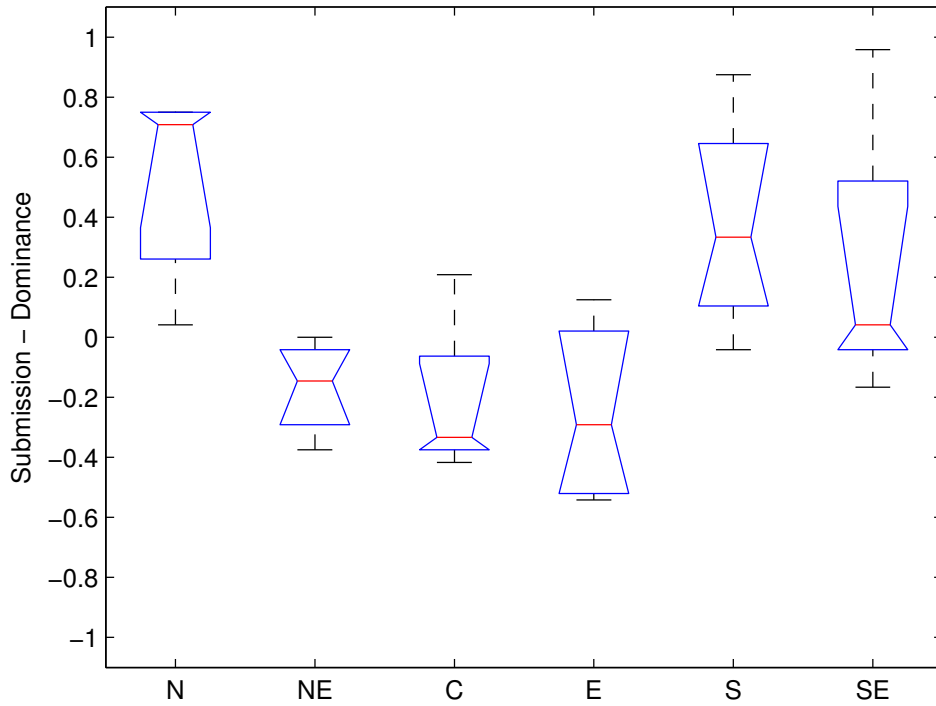


Figure 5.4: Dominance values for anger with directed head orientation and different gaze directions

following gaze directions: *North - Center* ($p < 0.05$) and *North - East* ($p < 0.05$) (see Figure 5.4). For Joy or Fear, we did not observe any significant differences in dominance perception for the chosen gaze directions.

Direct vs. Averted Gaze A two-tailed t -test on the dominance values assigned to Joy for direct and averted gaze while the head gaze was oriented towards the user revealed that the dominance value for averted gaze ($D = 0.35$) was significantly higher ($t = -3.0$, $df = 7.4$, $p < 0.05$) than the dominance value for direct gaze ($D = 0.77$). In contrast, the dominance value for Anger rose from -0.22 to 0.17 . However, this difference was not significant ($t = 2.2$, $df = 5.9$, $p = 0.07$). Fear, in comparison to Section 5.1.3 (*Direct vs. Averted Gaze*), also did not show a significant difference between direct and averted gaze, when the head orientation was directed towards the user ($t = 1.6$, $df = 5.0$, $p = 0.17$).

Influence of Head Orientation in Combination with Facial Displays on Dominance Ratings

Significant effects of different head orientations on dominance ratings were only obtained for Anger ($F(5, 113) = 2.9, p < 0.05, \eta^2 = 0.65$), Sadness ($F(5, 114) = 2.5, p < 0.05, \eta^2 = 0.44$) and Neutral ($F(5, 114) = 2.7, p < 0.05, \eta^2 = 0.33$) using a one-way ANOVA. A pairwise comparison for Anger within the groups with the post-hoc Tukey-HSD analysis revealed only significant differences between *Center* and *North East* ($p < 0.05$) head orientation (see Figure 5.5).

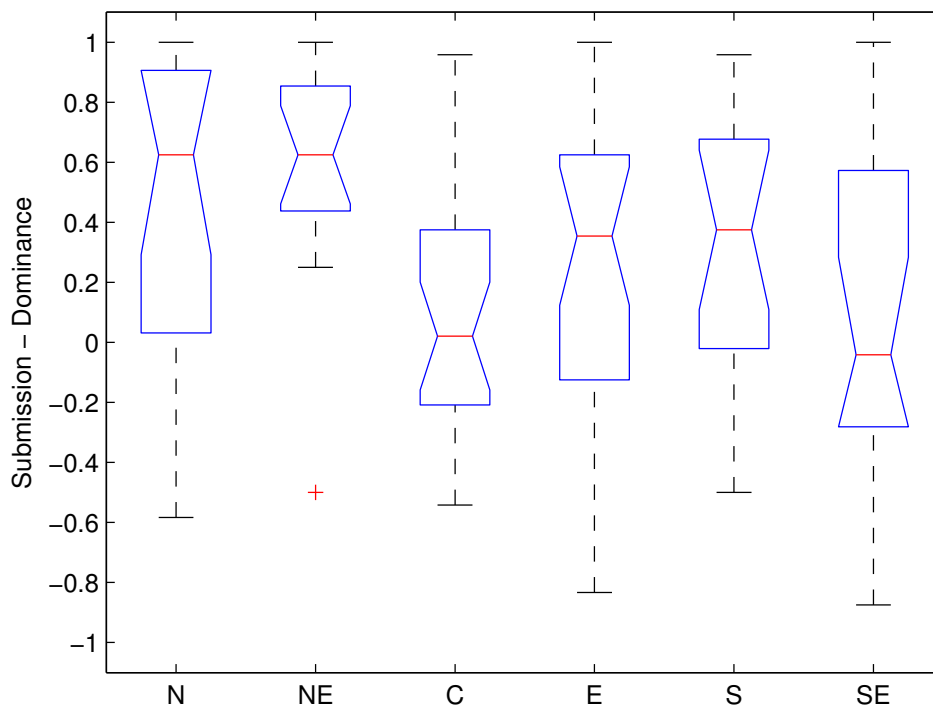


Figure 5.5: Dominance values for anger and head orientation

For Sadness, a post-hoc analysis did not reveal any significant differences with pairwise group comparisons. For Neutral, a Tukey-HSD post-hoc analysis showed significant group differences for *North - South* ($p < 0.05$) and *North - West* ($p < 0.05$) head orientation (see Figure 5.6).

Raised vs. Lowered Head For the six displayed emotions, we analyzed how a raised (*NW, N* and *NW*) in comparison to a lowered (*SW, S* and *SE*) head influenced the perception of dominance. A two-tailed t -test on a Neutral facial expression revealed significant differences between raised ($D = 0.41$) and lowered ($D = 0.20$) head orientations ($t = 2.30, df = 66.1, p < 0.01$). These findings are in line

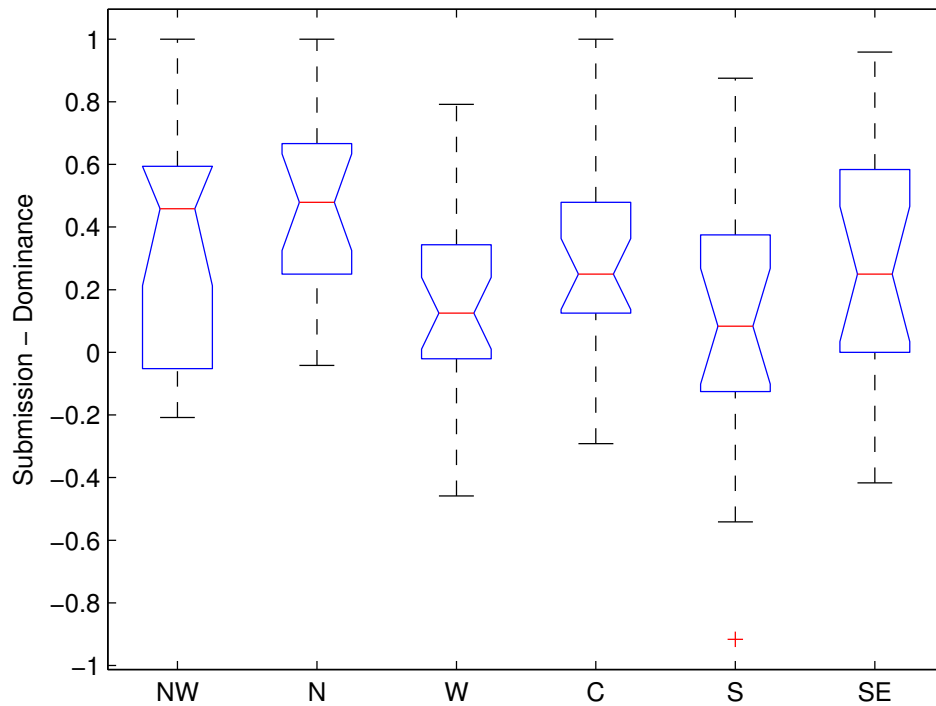


Figure 5.6: Dominance values for neutral and head orientation

with results from [Lance and Marsella \(2008\)](#), who analyzed perceived dominance for a virtual character without facial expressions.

In addition to Lance and Marsella, we found significant effects for angry and sad facial displays. In the case of Anger, the dominance value varied from 0.52 for a raised head to 0.22 for a lowered head ($t = 2.57$, $df = 64.0$, $p < 0.05$); in the case of Sadness it varied from 0.15 for a raised head to -0.16 for lowered head ($t = 3.26$, $df = 58.3$, $p < 0.01$). For the other investigated emotions, we did not find any significant differences between a bowed and a raised head.

That is, in the case of Anger, Sadness and Neutral, a raised head led to a higher dominance value compared to a lowered head.

Raised Head and Direct Gaze vs. Lowered Head and Averted Gaze We also investigated whether the effect obtained could be increased by showing a raised head in combination with direct gaze and a lowered head in combination with averted gaze. A significant difference in dominance ratings was observed for Anger and Disgust. A two-tailed t -test showed that the dominance value for Anger rose from 0.26 for a lowered head in combination with averted gaze to 0.76 for a raised head in combination with direct gaze ($t = 4.0$, $df = 29.0$, $p < 0.001$).

Also the dominance value for Disgust rose significantly from 0.27 in the case of a raised head and direct gaze to 0.54 in the case of a lowered head and averted gaze ($t = 2.5$, $df = 31.8$, $p < 0.05$). Joy could not be found to differ significantly ($p = 0.34$), just the same as Fear ($p = 0.38$) and Sadness ($p = 0.24$).

Perception Test

To check whether the chosen stimuli were perceived as they were meant, we conducted a small-scale study. We recruited eight subjects and presented them with screenshots of our virtual character (see Figure 5.1) showing the six emotions with the head and the eyes directed towards the subjects. The subjects had to decide which emotion word out of a list with six emotion words (joy, anger, sadness, fear, disgust and neutral) matched the currently shown facial expression best. Further, they were informed that one and the same emotion could be assigned to several faces and that they were allowed to name more than one emotion in case they had the impression that several emotions could fit.

	Joy	Anger	Sadness	Fear	Disgust	Neutral
Joy	8	–	–	–	–	–
Anger	–	5	1	–	3	1
Sadness	–	–	8	–	–	–
Fear	–	–	–	8	–	–
Disgust	–	4	–	–	7	–
Neutral	4	–	–	–	–	7

Table 5.4: Results of the perception test for emotional expressions of Alfred. Multiple nominations were allowed.

Table 5.4 presents the results of this test. The rows indicate onto which emotions the users mapped the shown facial expressions. For example, the display for disgust was recognized three times as anger and six times as disgust. To summarize the results, the displays for joy, sadness and fear were in all cases perceived as the emotion derived from the (Facial Expression Repertoire, 2008). Anger and disgust were a bit overlapping with each other. Neutral was perceived four times as joy, which is understandable when looking at Alfred’s neutral face (see Figure 5.1), which seems to be slightly smiling. Overall, the subjects were more or less able to assign the intended emotions onto the displayed facial expressions. As a consequence, we could exclude side effects resulting from badly displayed emotions.

5.2 Experiment II: Personality Perception – Extraversion, Agreeableness and Emotional Stability

Our premise is that personality types should be able to affect all possible facial actions directly and independently of the mood (Arya et al., 2006), or the emotions. Our intention is to explore which characteristics, or visual cues, are taken into account when perceiving personality. In this way more believable characters could be created.

Our second experiment is focused on the personality traits of *extraversion*, *agreeableness*, and *emotional stability*, taken from the Big-Five Factor model (Goldberg, 1992) and how they can be perceived when using two visual cues: head orientation and gaze. One of the main reasons to use this model is that it has already been widely used for the creation of virtual characters. Further, it defines each personality trait with a set of labels, producing a wide range of personality combinations.

In the following, we will present an experiment we conducted in order to investigate how head tilts and gaze of a virtual character influence the perception of extraversion, agreeableness, and emotional stability. The methodology consists of (1) creating different static images with combined head poses (up, center, down, side) and different gaze (up, center, down, side), (2) carrying out an online survey to measure the perception of personality on those images, and (3) evaluating the results to associate visual cues to extraversion, agreeableness, or emotional stability. We expect the following outcomes:

- The perception of the three personality traits Extraversion, Agreeableness, and Emotional Stability are not influenced whether the virtual character's head points to the left or to the right.
- The perception of Extraversion, Agreeableness, and Emotional Stability is influenced by the different directions where the head is pointing to. It makes a difference if the virtual character, for example, is looking to the upper-side corner or if the character is looking downwards to the middle.
- Dependent on the personality trait, direction plays a role in how these traits are perceived. We expect that it makes a difference, for example, how Extraversion is perceived in contrast to Agreeableness when the character is looking upwards.

- Not only head orientation influences the perception of the personality traits. Also variations of gaze directions further influence how the personality traits of the virtual character are perceived.

In the next section we describe the experimental method and how the virtual character was modeled, the questionnaire designed, and which were the characteristics of the participants. Subsequently, the results are presented and finally we conclude with a discussion of the study as well as the ideas and recommendations.

5.2.1 Experimental Study

For this study, our virtual agent named Alfred with slight background modification was used (see Figure 5.7 and Section 4.2).



Figure 5.7: The virtual character Alfred.

Stimuli

The visual cues, head movement and gaze orientation, targets for this study were calculated by varying horizontal and vertical angles, each in three symmetric steps. For both vertical and horizontal axis, the neutral center position remained at 0.0° . The positions for turning the head sideways were set as 8.5° or looking left, from the agent's point of view, and -8.5° for looking right. For tilting the head vertically, the high target was located at 8.0° degrees, whereas the low one was at -8.0° . Since the distance between neck and eye joints weakened the visible effect on the eye movement, the vertical angles had to be doubled for the eye targets.

All limits were chosen in regard to the goal that the pupils should remain visible, even when the eyes look in the opposite direction of the head.

By combining these angles, nine different targets could be provided for the survey. These were then converted to Cartesian coordinates using a fixed radius for all target angles, and sent to the virtual agent's inverse kinematics component (see Section 4.1.3) for every combination of head pose and gaze. The 81 resulting expressions were captured as screenshots.

However, to ensure a sufficient number of votes per picture, the number of samples had to be reduced and redundant combinations eliminated. To do this, previous observations were performed with a reduced group of users, obtaining as a result that the direction of lateral head movements would not cause much of a difference. Thus we decided to merge both *left* and *right* looking images into one "side" category. To keep the natural variation, about half of the required images were chosen randomly to either gaze in one or the other direction. The associated gaze targets were mirrored to keep the proper relation between head and eye movements.

In the end, from a set of 81 images, we worked with a reduced set of 54 (6 head directions \times 9 eye directions) different images of the Alfred character.

Measurement of Personality Traits

One of the most used trait models is the Big-Five Factor model, proposed by Goldberg (1992), which is based on five personality traits: extraversion, agreeableness, conscientiousness, emotional stability vs. neuroticism and openness. One of its advantages is its validity across cultures, which was proved for example by McCrae and Costa (1987) when validated the factors across six different cultures. Another widely respected trait model, based on factor analysis and a psychobiological basis is the Eysenck's model (Eysenck and Eysenck, 1975), composed by the traits






ACTION UNIT	DESCRIPTION	IMAGE
-----	head neutral	
AU 51	head turned left	
AU 52	head turned right	
AU 53	head up	
AU 54	head down	

Table 5.5: Varying head orientation.

of: psychoticism, neuroticism, and extraversion. Its advantage over the five-factor models is its stronger basis in physiological research, although it may not provide a good factor description of personality (Jackson, 2001). Based on these models, our aim with this research is to find which facial cues are characteristic for three personality traits: *extraversion*, *agreeableness*, and *emotional stability*.

According to Watson and Clark (1984), extraversion is defined by seven components: venturesomeness, affiliation, positive affectivity, energy, ascendance, and ambition. Nevertheless, McCrae and John (1992) divide “affiliation” in warmth and gregariousness. On the other hand, people low in extraversion are described as quiet, reserved, retiring, shy, silent, and withdrawn. However, different psychologists varied the list of components for each trait.

Neuroticism, or low emotional stability, represents individual differences in the

tendency to experience distress, and in the cognitive and behavioral styles that follow from this tendency. In addition, individuals low in neuroticism are not necessarily high in positive mental health, however that may be defined as they are simply calm, relaxed, even-tempered, and unflappable.

The agreeableness factor was taken into consideration because the idea is to generate agents that the user can interact with, and this factor measures the level of friendliness, cooperation, generosity, among other socially, or human related characteristics.

Questionnaire

133 subjects (47 female and 86 male) participated in the experiment through an online questionnaire. The mean age was 26.6 ($SD = 8.8$). The questions were provided in English, as well as their validated translations into German or Spanish. The questionnaire consisted of 54 static images, where each image was judged at least 10 times. The images corresponded to a virtual character in which head orientation (head up front, head up side (side-indifferent), head front, head side (side-indifferent), head down front, head down side (side-indifferent)) and gaze (gaze up front, gaze down front, gaze up side (side-indifferent), gaze down side (side-indifferent)) were combined.

Then the experimental stimuli which consisted of 15 images per user were presented one at a time, in random order. For each stimulus the participant had to answer to six items of the Ten-Item Personality Inventory (TIPI) (Gosling et al., 2003) presented in a 7-point Likert Scale, where 1 corresponded to “disagree strongly” and 7 to “agree strongly”. Table 5.6 shows the items presented for each image of the questionnaire (see Appendix C.3).

TRAIT	ITEM
Extraversion	Extraverted, enthusiastic
	Reserved, quiet
Agreeableness	Critical, quarrelsome
	Sympathetic, warm
Emotional Stability	Anxious, easily upset
	Calm, emotionally stable

Table 5.6: Questionnaire items for perception of head orientation.

5.2.2 Results

In the following, we will present an experiment we conducted in order to investigate how gaze and head tilts of a virtual character influence the perception of extraversion, agreeableness, and emotional stability. The virtual agent Alfred is over all ratings neither perceived as extraverted nor as introverted ($M = 3.7, SD = 1.2$). Further, he is perceived as neutral regarding agreeableness ($M = 3.8, SD = 1.4$). However, the subjects perceived Alfred as slightly emotional stable ($M = 4.3, SD = 1.4$).

Looking to the Right and to the Left

To study if the side the agent is looking to has an influence on the perception of personality, we assumed that in general there are *no noticeable difference* among the personality traits whether the agent looks to the right or to the left. The method was to apply a two-tailed independent t -test to the overall values for extraversion, agreeableness, and emotional stability dependent on which side the virtual agent is looking.

The results of the t -test with 329 degrees of freedom for the trait of extraversion, $t(329) = -1.2, p = 0.25, r = 0.07$, showed that there was no significant difference for extraversion between Alfred looking to the left ($M = 3.8, SD = 1.2$) and looking to the right ($M = 3.9, SD = 1.3$), given that the obtained p -value is above the probability threshold. Moreover, the effect size r has a value below .1, which demonstrates the weak relationship between extraversion and the side to where the character is looking to. Therefore, we can accept the assumption of no noticeable difference in extraversion when the agent looks to the left or to the right.

Agreeableness also did not show any significant differences for the virtual character Alfred between looking to the left ($M = 3.8, SD = 1.5$) and looking to the right ($M = 3.7, SD = 1.4$), $t(329) = 0.62, p = 0.54, r = 0.03$. Again, given that the obtained p -value is above the threshold and the effect size r is below .1, a weak relationship between agreeableness and the side to where the character is looking to is shown. Hence, we also can accept the assumption of no noticeable difference.

Finally, emotional stability as well did not show any significant differences between looking to the left ($M = 4.4, SD = 1.4$) and looking to the right ($M = 4.4, SD = 1.2$), $t(329) = -0.35, p = 0.73, r = 0.02$. In the latter case, as with extraversion and agreeableness, the effect size r is for all three personality dimensions below .1 and thus can be further interpreted as not even a small effect.

Extraversion

Positioning the head to the upper side got the highest rating for extraversion, while redirecting the head to the lower middle got the lowest rating for extraversion (see Figure 5.8 and Figure 5.9).

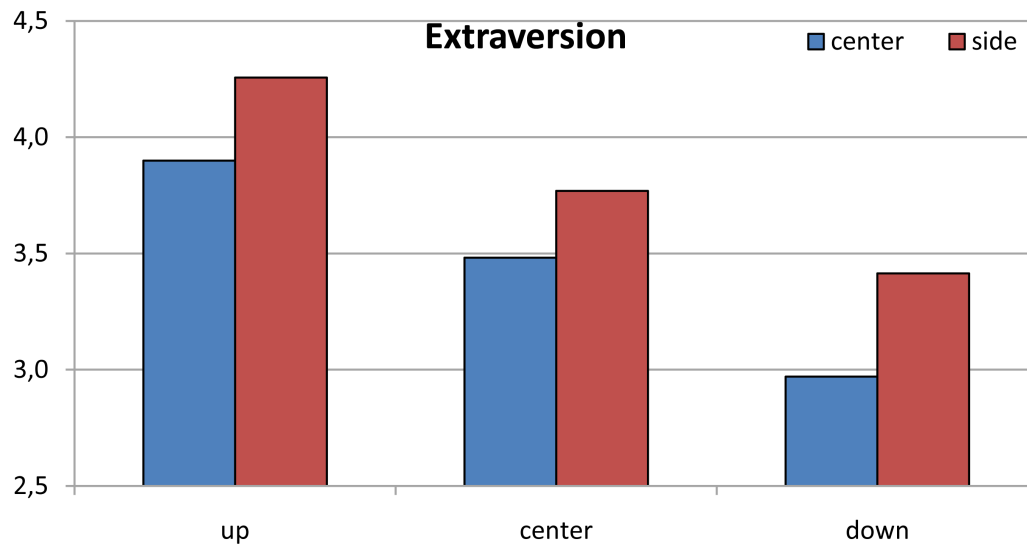


Figure 5.8: Mean values for *extraversion* dependent on the head orientation.

The one-way ANOVA showed that there was a significant effect on the perception of Extraversion on levels of the different head orientations, $F(5, 663) = 15.4$, $p < 0.001$, $\omega^2 = 0.10$. Tukey post hoc tests revealed several significant differences within the perception of extraversion dependent on the head orientation (see Table 5.7 where each row and column corresponds to the combination of vertical and horizontal positioning, e.g. U-C means “upper-center”).

The virtual character with its head pointing to the upper side ($M = 4.3$, $SD = 1.2$) was perceived significantly less extraverted than the heads pointing to the center side ($M = 3.8$, $SD = 1.3$, $p < 0.05$), to the center middle ($M = 3.5$, $SD = 1.2$, $p < 0.001$), to downwards side ($M = 3.4$, $SD = 1.1$, $p < 0.001$), and to downwards middle ($M = 3.0$, $SD = 1.1$, $p < 0.001$).

An upper middle head position ($M = 3.9$, $SD = 1.2$) is perceived as less extraverted than the heads looking to the center ($M = 3.5$, $SD = 1.2$, $p < 0.1$), looking to the lower side ($M = 3.4$, $SD = 1.1$, $p < 0.05$), and to the lower middle ($M = 3.0$, $SD = 1.1$, $p < 0.001$).

A head directed to the center side ($M = 3.8$, $SD = 1.3$) is perceived as less extraverted than a head looking downwards to the middle ($M = 3.5$, $SD = 1.2$, $p < 0.001$).

As we applied a two-tailed post hoc test, the significant results are also valid vice versa.

	U - S	U - M	C - S	C - M	D - S	D - M
U - S	—	n.s.	*	***	***	***
U - M	n.s.	—	n.s.	+	*	***
C - S	*	n.s.	—	n.s.	n.s.	***
C - M	***	+	n.s.	—	n.s.	*
D - S	***	*	n.s.	n.s.	—	+
D - M	***	***	***	*	+	—

Table 5.7: Post-hoc comparisons for *Extraversion* and the varying head orientations (U = up, C = center, D = down, S = side, M = middle, $^+p < 0.1$, $^*p < 0.05$, $^{***}p < 0.001$, n.s. = not significant).

In general we can see that the virtual character averting its head gaze to the side increases the perception of extraversion independent if the virtual character looks upwards, to the middle or downwards (see Figure 5.8). And in general there seems to be trend dependent on the vertical head orientation too. A raised head is perceived as more extraverted than a head oriented to the middle and head looking downwards.



Figure 5.9: The head orientation (*center down*) with the lowest rating (left) and the one (*up side*) with the highest rating (right) for *Extraversion*.

We could not find any significant differences among the six head orientations within the nine eye directions.

Agreeableness

The highest value for agreeableness was achieved when the virtual character looked to the lower middle. The lowest value was achieved for a virtual character looking to the upper middle (see Figure 5.10 and Figure 5.11).

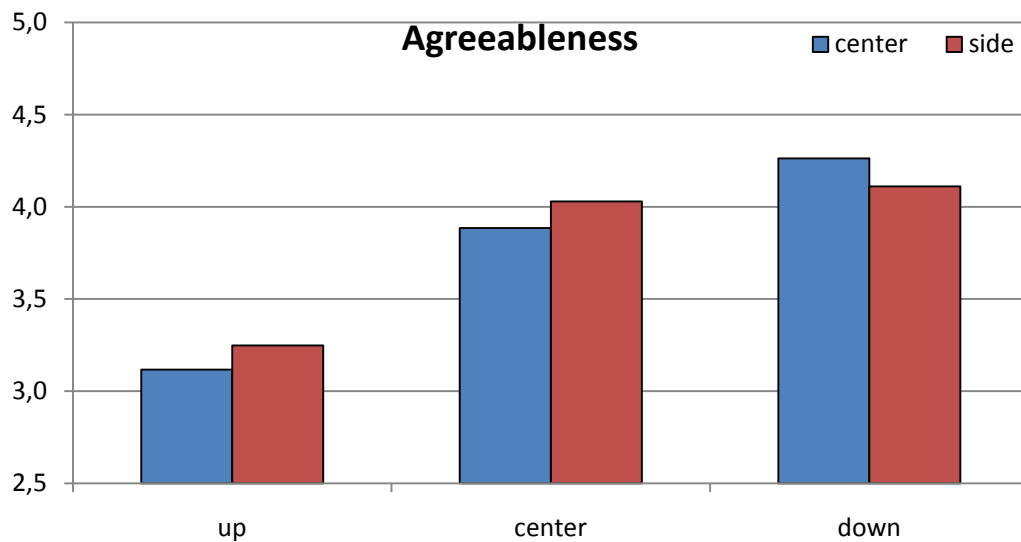


Figure 5.10: Mean values for *agreeableness* dependent on the head orientation.

There was a significant effect on the perception of agreeableness on levels of the different head orientations, $F(5, 663) = 14.4, p < 0.001, \omega^2 = 0.09$. Tukey post hoc tests revealed several significant differences within the perception of Agreeableness dependent on the head orientation (see Table 5.8).

The head directed to the upper side ($M = 3.3, SD = 1.3$) was perceived as less agreeable than a head directed to the center side ($M = 4.0, SD = 1.4, p < 0.001$), the center ($M = 3.9, SD = 1.3, p < 0.01$), the lower side ($M = 4.1, SD = 1.4, p < 0.001$), and the lower middle ($M = 4.3, SD = 1.2, p < 0.001$).

The virtual character with the head looking to the upper center ($M = 3.1, SD = 1.3$) was perceived as less agreeable than a head looking to the middle side ($M = 4.0, SD = 1.4, p < 0.001$), looking to the center ($M = 3.9, SD = 1.3, p < 0.001$), looking downwards to the side ($M = 4.1, SD = 1.4, p < 0.001$), and looking downwards to the middle ($M = 4.3, SD = 1.2, p < 0.001$).

The lowest values for agreeableness were achieved for the virtual character looking upwards. Higher values could be achieved for Alfred looking to the middle and even slightly higher values were achieved for looking downwards (Figure 5.10).

	U - S	U - M	C - S	C - M	D - S	D - M
U - S	—	n.s.	***	**	***	***
U - M	n.s.	—	***	***	***	***
C - S	***	***	—	n.s.	n.s.	n.s.
C - M	**	***	n.s.	—	n.s.	n.s.
D - S	***	***	n.s.	n.s.	—	n.s.
D - M	***	***	n.s.	n.s.	n.s.	—

Table 5.8: Post-hoc comparisons for *agreeableness* and varying head orientation (U = up, C = center, D = down, S = side, M = middle, ** $p < 0.01$, *** $p < 0.001$, n.s. = not significant)



Figure 5.11: The head orientation (*center up*) with the lowest rating (left) and the one (*center down*) with the highest rating (right) for *Agreeableness*.

Also for agreeableness we could not find any significant differences for the varying gaze directions dependent on the six head orientations.

Emotional Stability

The virtual character looking to the middle side achieved the highest ratings for emotional stability and the virtual character looking to the upper middle achieved the lowest ratings (see Figure 5.12 and Figure 5.13).

There was a significant effect on the perception of emotional stability on levels of the different head orientations, $F(5, 663) = 3.6, p < 0.01, \omega^2 = 0.02$.

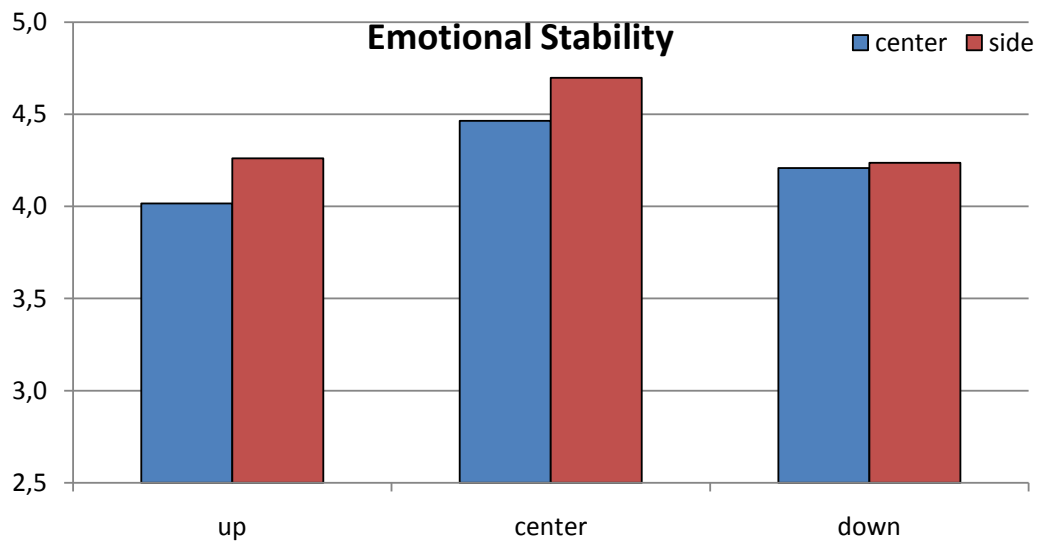


Figure 5.12: Mean values for *Emotional Stability* dependent on the head orientation.

Tukey post hoc tests revealed only one significant difference within the perception of emotional stability dependent on the head orientation.

The virtual character directing its head to the center side ($M = 4.7$, $SD = 1.2$) was perceived with significantly lower emotional stability than the virtual character looking to the upper middle ($M = 4.0$, $SD = 1.3$) with $p < 0.001$.

Looking at middle side ($M = 4.7$, $SD = 1.2$) was perceived as more emotional stable than looking to the lower middle ($M = 4.2$, $SD = 1.5$, $p < 0.1$).

For emotional stability the highest value were perceived when the virtual character's vertical head orientation was directed to the middle, independently of the looking side. And further, looking upwards or downwards both got in general the lowest values for emotional stability (see Figure 5.12).

5.3 Conclusion

In this chapter we investigated, first, how the interplay of facial display, gaze and head tilts determines the perception of dominance. We could show that dominance ratings are influenced by emotional facial expressions. Higher dominance values were found for facial expressions conveying joy, anger and disgust. The dominance rating for a neutral facial expression, however, was significantly lower than that for joy, anger or disgust. Sadness and fear were perceived significantly less dominant in our experiment than joy, anger, disgust and neutral.



Figure 5.13: The head orientation (*center up*) with the lowest rating (left) and the one (*center side*) with the highest rating (right) for *emotional stability*.

Further, we analyzed how dominance perception changes with varying gaze direction and head orientation. Our hypothesis that an averted gaze increases the degree of perceived submissiveness could not be confirmed in general. Only joy was perceived as less dominant when the gaze was averted. In contrast, anger and fear led to an increase in dominance in combination with averted gaze. Further, we found that gaze aversion had no influence on dominance ratings in combination with faces showing sadness, disgust or a neutral expression. Significant differences between an upward and downward directed head orientation could only be found for a neutral state, anger and sadness. Here, a lowered head orientation reduced the perception of dominance.

Finally, we could show that an upward head orientation in combination with direct gaze was rated as significantly more dominant than a downward oriented head with averted gaze direction for anger and disgust. To summarize these findings, it matters where a virtual agent directs its attention dependent on its current affective state, and such effects need to be taken into account when modeling attentive affective agents.

In the second part, we investigated how the gaze and head orientation determines the perception of extraversion, agreeableness and emotional stability. The obtained results brought light on the question if certain visual cues could be associated with personality traits. With the experiment we found that, for the Alfred character the “up-side” head orientation is related to extraversion, “center-down” head orientation is related to agreeableness, and “center-side” head orientation is related to emotional stability. We also confirmed our hypothesis that head side

orientation, i.e. if the character's head or gaze is oriented to the left or to the right, does not influence the perception of personality traits.

An important aspect of this work is the possibility to obtain facial/head visual cues for certain personality traits that have been not studied before, as emotional stability and agreeableness. This opens a very beneficial field not only in the generation of personality-based characters, but on the recognition of personality traits. One of the applications of this investigation would be in head trackers where the personality of the subject might be depicted based on his/her head orientation.

From our results we could also observe that people take into consideration more than a limited number of visual cues to infer personality. Personality is not only influenced by head orientation or gaze. Take an "up-side" head and combine it with a facial expression of sadness, and the perception of personality could be other than extraversion.

Chapter 6

Study III: Gaze Interaction with Virtual Characters

In this chapter we will present an interactive gaze model for virtual characters which focuses on the synchronization between the virtual character's gaze and the current user's gaze. We will first show how the eye orientation of a virtual character influences the perception of the "being-seen" of a human. Next, we will describe how a generic gaze model looks like that takes the user's gaze into account. And in the following sections we will first evaluate the non-verbal interaction with the interactive gaze model and second the verbal interaction with the interactive gaze model.

The evaluation of the gaze interaction with a non-verbal scenario will make use of a flirt scenario. In human-human conversation, the first impression decides whether two people feel attracted by each other and whether contact between them will be continued or not. Starting from psychological work on flirting, we implemented an eye-gaze based model of interaction to investigate whether flirting tactics help improve first encounters between a human and an agent. Unlike earlier work, we concentrate on a very early phase of human-agent conversation (the initiation of contact) and investigate which non-verbal signals an agent should convey in order to create a favorable atmosphere for subsequent interactions and increase the user's willingness to engage in an interaction with the agent. To validate our approach, we created a scenario with a realistic 3D agent called Alfred that seeks contact with a human user. Depending on whether the user signals interest in the agent by means of his or her gaze, the agent will finally engage in a conversation or not (Bee et al., 2009a).

The evaluation of the gaze model with a verbal interaction will make use of an interactive storytelling scenario. We investigate the user's gaze behavior during

the conversation with an interactive storytelling application. We present an interactive gaze model for embodied conversational agents in order to improve the experience of users participating in Interactive Storytelling. The underlying narrative in which the approach was tested is based on a classical XIXth century psychological novel: *Madame Bovary*, by Flaubert. At various stages of the narrative, the user can address the main character or respond to her using free-style spoken natural language input, impersonating her lover. An eye tracker was connected to enable the interactive gaze model to respond to user's current gaze (i.e. looking into the virtual character's eyes or not). We conducted a study with 19 students where we compared our interactive gaze model with a non-interactive gaze model that was informed by studies of human gaze behaviors, but had no information on where the user was looking. The interactive model achieved a higher score for user ratings than the non-interactive model. In addition we analyzed the users' gaze behavior during the conversation with the virtual character (Bee et al., 2010b).

6.1 Eye Orientation in Human-Agent Interaction

We conducted a study to investigate whether the gaze focus of a virtual character gives the users the impression of seeing through them (see Figure 6.1) or not and which eye orientation is the right one not to be seen through. We designed five different focusing points of the virtual characters eyes with different distances from the user. Two focusing points in front of the user (1 and 2), one focusing directly the head of the user (3), one focusing point behind the user (4) and one with parallel gaze orientation (5) which corresponds to gazing at infinity. Eleven subjects had to rate the five different focusing points sitting in front of the computer screen and a virtual character's head directed straight to them. The order of the samples was randomized and in total the subjects had to rate each focusing point three times on a 5-point scale (values from 0 to 4). We asked two questions about the focus perception of the virtual character. The first question (F1) was if the users had the feeling that the character is looking at them like a real person and the second question (F2) was about the feeling if the character is seeing through them.

To analyze differences between the different eye focus angles, we use a one-way ANOVA and the Tukey-HSD for the post-hoc two-sided pairwise comparison. The ANOVA test reveals that users significantly perceive differences if the user is focusing on them or not. Question F1, asking if the character is focusing like a real person, shows significant differences between the five varying eye focus points ($F(4, 160) = 16.8, p < 0.001$). The Tukey-HSD analysis reveals that gaze focus (1) significantly differs from all other focus points with $p < 0.001$ to eye angle (2), (3),

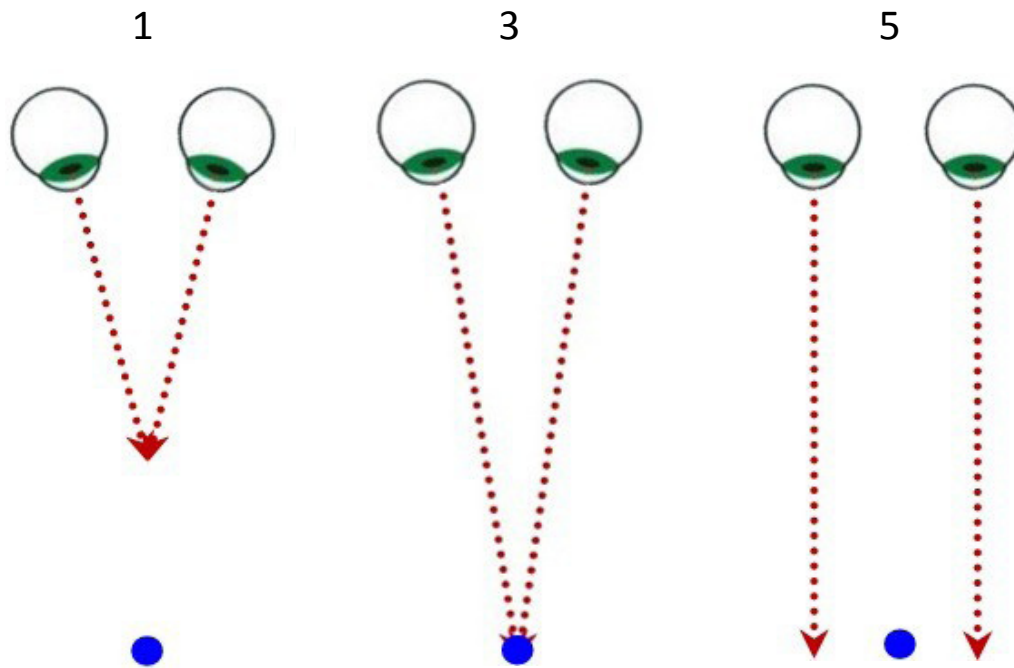


Figure 6.1: Variations of the eye angle: (1) focusing a point in front of the subject, (3) focusing the subject, (5) parallel eye direction.

(4) and $p < 0.01$ to (5) (see Figure 6.2, left chart). Also for question F2, about the character seeing through the user, the ANOVA test reveals significant differences ($F(4, 160) = 10.5, p < 0.001$). The Tukey-HSD analysis reveals that the eye focus point (1) only significantly differs from focus point (3) ($p < 0.01$) and focus point (5), the one that was parallel, significantly differs from focus point (2) and (3) with $p < 0.001$ and from point (4) with $p < 0.01$.

We found that the virtual character's eye orientation matters regarding the "being-seen" perception. This means, if a virtual character focuses a spot in front of the user (1) or looks parallel (5) the user perceives this as unnatural and not being focused. Thus the eye orientation needs to be adjusted to the current user-agent distance.

6.2 Interactive Gaze Model for Human-Agent Interaction

In this section we start from the gaze model developed by Fukayama and colleagues which allows us to specify a number of gaze parameters that influence the impression a character conveys (Fukayama et al., 2002). Their model includes

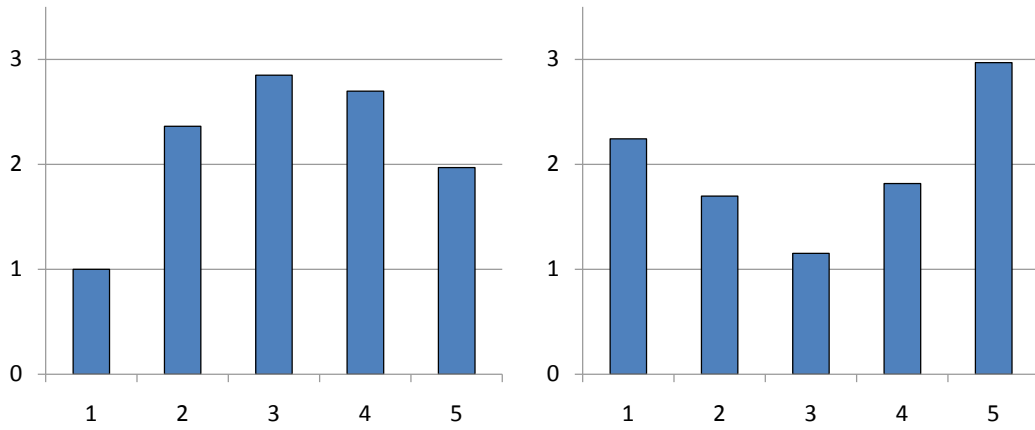


Figure 6.2: The charts show the results for the focus experiment (left: looking like a real person, right: seeing through me).

two states: looking at the user and averting the gaze from the user. Three parameters define how often, how long (500 to 2000 ms) and where the virtual agent looks. The gaze targets consist of a set of random points from either all over the scene, above, below or close to the user. The probabilities of changing from one state to the other or staying in the same state depend on the amount and the mean duration of the gaze parameters. Fukayama and colleagues rated the impression particular gaze patterns conveyed, that were produced by modifying the gaze parameters. They found that a medium amount of gaze and a mean duration between 500 to 1000 ms conveys a *friendly* gaze behavior. The orientation of the gaze direction did not play a decisive role in distinguishing between friendly and dominant gaze behavior, except a downward gaze was considered as less dominant. Fukayama and colleagues evaluated their gaze behavior model by only displaying eyes to the users. Thus, we evaluated their model with a full virtual head that in addition moves his head and eyes. Basically, we followed their settings, but distinguished whether the agent is speaking or listening.

Our gaze model (see Figure 6.3) was extended with further parameters as our virtual agent is capable of reacting to the user's current gaze using an eye tracker. The maximal and minimal duration of mutual gaze can now be set as well. Furthermore, we may indicate the maximal duration the virtual agent gazes around. In the Gaze averted state the virtual agent does not look at the user. The virtual agent looks randomly at some predefined points besides the user's head. The probability for the states changes (i.e. P_{UU} , P_{AA} , P_{AU} and P_{UA}) can be individually adjusted. In the Gaze at user state the virtual agent looks at the user. Whenever the system detects mutual gaze [MG] between the user and the virtual character the system changes from Gaze at user to Gaze averted after a specific defined time (e.g. 1 ± 0.5 seconds) expires. The time interval of the state machine can be

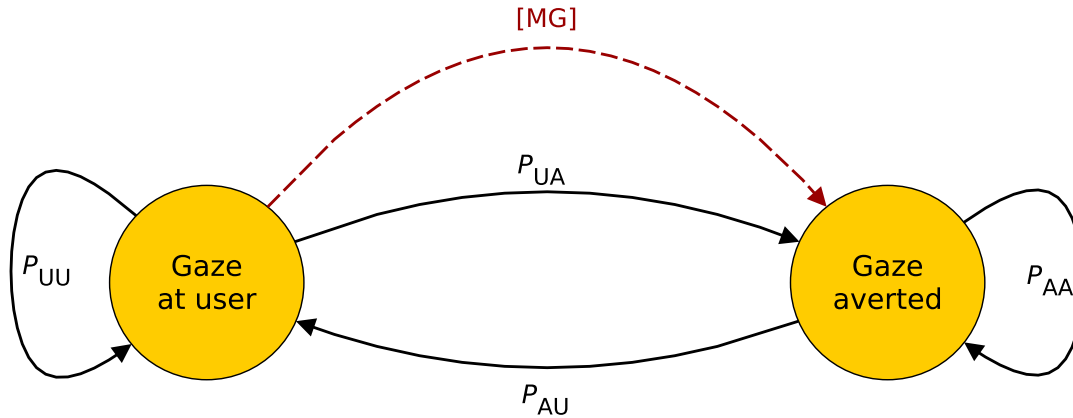


Figure 6.3: The gaze model reacts on mutual gaze [MG] between user and virtual character. If a mutual gaze is recognized, the state switches from Gaze at user to Gaze averted for a specific time.

configured as well.

6.3 Experiment I: Non-Verbal Interaction

In human-human conversation, the first impression decides whether two people feel attracted by each other and whether contact between them will be continued or not. A large industry has developed around the production of training material that intends to give advice to people that wish to present themselves in a positive light to new people (for example, see (Cohen, 1992), for a guidebook on flirting). A recommendation that can be found in almost any guidebook that prepares people for such first encounters, whether be it in a date or a job interview, is to show a genuine interest in the conversational partner using gaze and smiles. In this way, instant rapport is built up which creates a good starting point for subsequent interactions (see (Argyle and Cook, 1976) and (Kleinke, 1986)).

There has been a significant amount of work on embodied conversational agents that make use of non-verbal behaviors to establish rapport with a human user. There is empirical evidence that rapport-building tactics also work for human-agent communication, see, for example, (Gratch et al., 2006). Most work focuses on the use of non-verbal signals during a dialogue taking it for granted that the human has an interest in communicating with the agent. This experiment concentrates on a very early phase of human-agent communication (the initiation of contact) and investigate which non-verbal signals an agent should convey in order to create a favorable atmosphere for subsequent interactions and increase the user's willingness to engage in an interaction with the agent. In particular, we are

interested in the question of whether it is possible to give a human user the feeling that an agent has a genuine interest in him or her. We consider work on flirt tactics as a useful resource to implement agents that have these capabilities.

To validate our approach, we created a scenario with the realistic 3D agent called Alfred that was introduced in Chapter 4.2 that seeks contact with a human user. Depending on whether the user signals interest in the agent by means of his or her gaze, the agent will finally engage in a conversation or not. In addition, we employ a contact-free eye tracker that provides us continuously with information on the user's gaze. In order to use means to analyze and generate social signals effectively, the agent has to sense the user's gaze and to align it in parallel with its own behaviors.

6.3.1 Gaze Model for Human-Agent Interaction

As a basis for our research, we rely on the approach by Givens (1978) who distinguishes between five phases of flirting:

- The *attention phase* describes the phase in which men and women arouse each other's attention. It is characterized by ambivalent non-verbal behavior, such as a brief period of mutual gaze broken by downward eye aversion, reflecting the uncertainty of the first seconds.
- In the *recognition phase*, one person recognizes the interest of the other. He or she may then discourage the other person, for example, by a downward gaze, or signal readiness to continue the interaction, for example, by a friendly smile.
- After mutual interest has been established, the man or woman may be initiated to the *interaction phase* and engage in a conversation.
- In the *sexual-arousal* and *resolution phases*, the relationship between man and woman intensifies. These two phases are not further described here because of their missing relevance to human-agent communication.

Our system will cover the attention, recognition and the initiation of the interaction phase. In addition to the work by Givens, our gaze based interaction system incorporates findings from Tramitz (1992) and Bossi (1995). In particular, we rely on their work to determine the timing of gazes. Tramitz (1992) analyzed the flirt behavior in a study with 160 school students. She found that the initiation of a first encounter decides on the continuation of the flirt interaction. Bossi (1995)

used the same study but analyzed couples with different levels of interest on each other. Flirting couples seemed to use more time, up to three times, to gaze at each other. Further, the first gaze and following single gazes lasted longer.

Attention Phase

The implementation of the attention phase (see Figure 6.4) is motivated by typical behavior sequences described in (Givens, 1978). The attention phase starts at the point when the human and the virtual agent take notice of each other. The virtual agent shows a slightly friendly facial expression and the gaze is averted from the user. First, the agent's eyes only wander around the room for a while until they meet the user's eyes. After that, the virtual agent will engage in an interplay of mutual and averted gaze.

While the virtual agent gazes randomly around the room, the system checks whether the user gazes at the agent. If this is the case, the agent establishes gaze contact with the user. The system is now in the hold-mutual-gaze-state and the agent tries to hold eye contact until a specified time interval elapses. It then breaks off eye contact by a downward gaze accompanied by a smile. To avoid that unconscious very brief sweep gaze behaviors of the user are by mistake categorized as mutual gaze, eye contact with the user will be taken into account only after a certain duration. In case the user breaks off eye contact before the maximal time interval elapses, the agent will look away as well, however, without showing a smile since the user would not recognize the facial expression anymore. If the user does not respond to eye contact established by the agent, the agent will avert its gaze again. After each successful or failed eye contact, the system will return to the state of looking around trying to establish eye contact again or to respond to the user's gaze.

This loop is repeated until one of the following terminating conditions is fulfilled. In the positive case, a certain number of mutual gazes could be established and the system transits to the next phase. In the negative case, the agent has attempted to establish gaze contact with user in vain and breaks off the complete interaction due to missing interest of the human flirt partner. After each successful gaze contact, the emotional state of the agent improves and its facial display becomes more joyful. After each failed attempt to establish gaze contact, the emotional state of the agent becomes worse and it looks more sad.

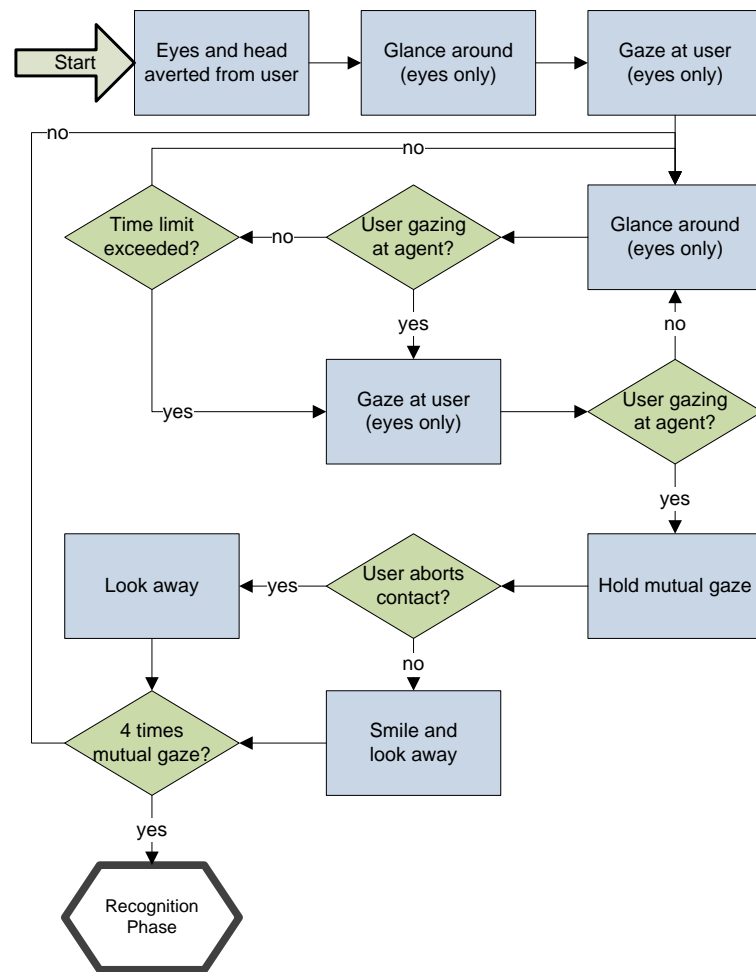


Figure 6.4: The *attention phase* models the phase in which men and women arise each other's attention. It is characterized by ambivalent non-verbal behavior, such as a brief period of mutual gaze broken by downward eye aversion, reflecting the uncertainty of the first seconds.

Recognition Phase

Similar to the attention, the recognition phase (see Figure 6.5) is based on the interplay between mutual and averted gaze behavior, save that the durations of the mutual gazes increase. Further, the virtual agent smiles more often and uses more distinct flirt signals (i.e. eyebrow flash, pout or raise of the upper eyelid). Since the agent's self confidence has increased after the successful attention phase, it tries to establish eye contact more often and eye movements are supported by head movements to show a more obvious interest in the other person. Just as the attention phase, the recognition phase can still fail. Namely, in case the agent unsuccessfully tried to establish mutual gaze for several times. Or, if a particular

number of mutual gazes has been set up, the recognition phase was successful and completed. This will lead to the next phase, which is the interaction phase.

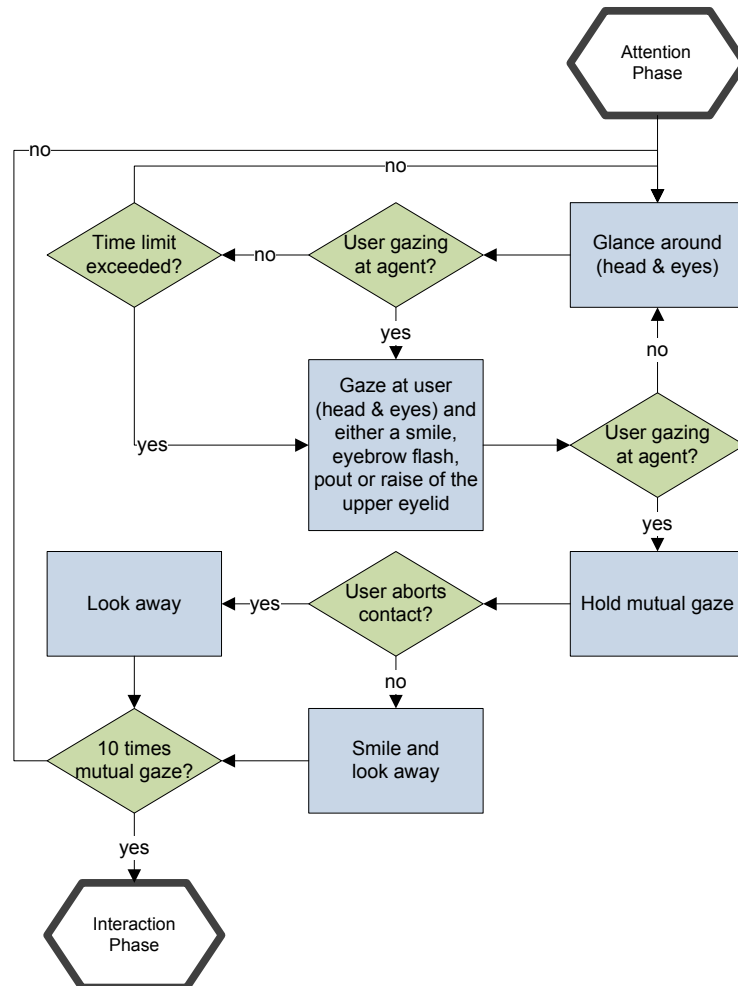


Figure 6.5: In the *recognition phase*, one person recognizes the interest of the other. He or she may then discourage the other person, for example, by a downward gaze, or signal readiness to continue the interaction, for example, by a friendly smile.

Interaction Phase

After successfully completing the two previous phases, verbal communication will be initiated in the interaction phase. The virtual agent verbally addresses the user using small talk strategies adopted from [Tramitz \(1992\)](#).

6.3.2 System

The system for gaze based interaction between the virtual agent and human consists of an eye tracker, our virtual character Alfred (see Figure 6.6), which was introduced in Chapter 4.2, and the program logic for the interaction. We use a contact-free eye tracker (SMI iView X RED) which allows the user to move relatively free.



Figure 6.6: The virtual character Alfred with a skyline in the background.

System for Interaction

To be able to detect where the user is looking at, we connected the eye tracker with the virtual 3D-world. Ray casting allows us to map the screen coordinates obtained from the eye tracker to the objects in the virtual world. In this vein, we are able to detect whether the user looks at the virtual agent, the left eye or the right eye or something else in the virtual scene. This was necessary for the gaze based interaction on a level of mutual gaze and to see, if the user is looking at Alfred or not.

The flirt behavior can be varied by parameters. A confidence value defines the agent's level of self assurance and influences the probability that the agent initiates

gaze interaction. The maximal and minimal duration of mutual gaze can be set as well. Furthermore, we may indicate the maximal duration the virtual agent gazes around. For the attention and recognition phases, the maximum number of trials to initiate mutual gaze before it fails can be defined. Finally, we may specify how long the virtual agent waits until the user responds with mutual gaze. These parameters are stored in a XML file and can be easily adjusted.

The virtual agent is able to direct his gaze using his eyes only or his head and his eyes in combination.

Flirt signals are displayed whenever mutual gaze occurs. In the attention phase, flirt signals are rarely sent whereas in the recognition phase, flirt signals are an integral part of the interaction. Whenever mutual gaze occurs, the virtual agent sends with a probability of $1/3$ one of the following flirt signals: an eyebrow flash, a pout, a raise of the upper eyelid or a smile.

The agent's mood changes dependent on the number of successful mutual gazes. The more mutual gazes occur, the friendlier Alfred's facial expression becomes. Vice versa, the agent's happiness declines if there is no reaction from the user to an attempt to establish mutual gaze.

6.3.3 Evaluation

We conducted an empirical study to demonstrate the benefits of gaze based interaction in combination with flirting tactics. Our main focus was to figure out if the users realize the virtual agent's interest. Further, we would like to investigate the impact of an gaze based interaction system on user engagement. Finally, we were interested in finding out whether a gaze based interaction system works with a life-size setting. Thus, the study focuses on the following points:

1. The flirting agent is able to show the user that it has an interest in him through its gaze and facial expression behavior, and the user will perceive this behavior as flirting.
2. The integration of flirting tactics has a positive impact on the perception of the agent and the interaction with him and thus contributes to the user's engagement.
3. By tracking the user's gaze and responding to it in real-time, the effects can be increased.

Setting

The optimal dimensions for such a video projector based eye tracking setting are limited. The user is placed in front of a table on which the eye tracker was placed. The eye tracker with an incline of 23° is placed 80 cm above ground and 140 cm away from the projection surface. The user is seated 60 - 80 cm in front of the eye tracker. In total the user is about 2 m away from the virtual agent, which is within the *social space* according to (Hall, 1963). The projection surface sizes 120×90 cm, which displays the virtual agent in life-size (see Figure 6.7).

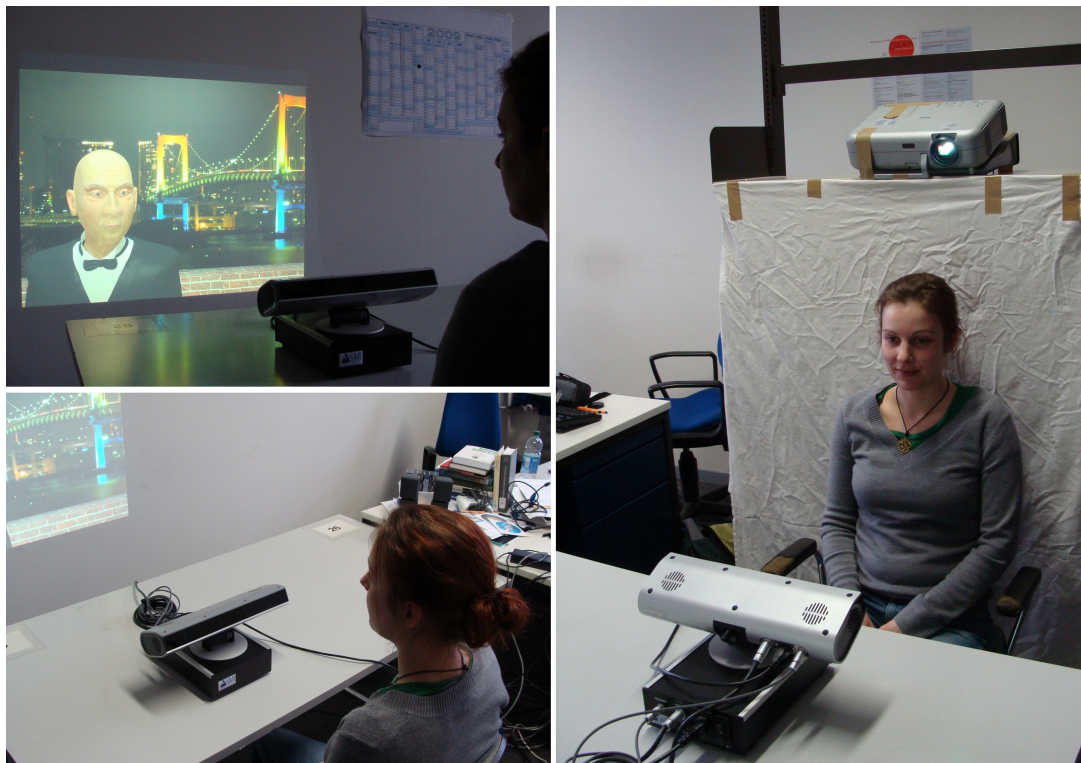


Figure 6.7: Set-up for the gaze based interaction application from different perspectives.

To avoid that the user automatically stares at the virtual agent (which would happen if it was placed in the center of the visual display), we placed it on the left side. To offer an enriched scene where the user has the choice to look away from the virtual agent, we added a city skyline (see Figure 6.6).

Interaction Modes Apart from the fully interactive version where the agent recognizes and responds to the user's gaze behavior in real-time, two further gaze behavior variants were created to demonstrate the benefits of the gaze behavior model described in Section 6.3.1: a non-interactive version which implements an

ideal flirt behavior derived from the literature and a non-interactive version which implements an *anti-flirt* behavior.

In the non-interactive ideal version the virtual agent behaves like in the interactive version except for that it does not respond to the user's gaze behavior, but assumes a perfect gaze behavior from the user and thus follows a fixed sequence. The gaze directions while glancing around the scene are still randomly selected, but the virtual agent gazes at the user always with the same duration, no matter whether the user returns the gaze or not.

In the non-interactive anti-flirt version, by contrast, the virtual agent behaves contrarily to the typical flirt behaviors previously described. The duration of the mutual gaze is increased from 3 seconds to 7 seconds, which is commonly considered as staring. Furthermore, the facial expression remains neutral, which can be interpreted as a bored attitude towards the user. Finally the virtual agent looks away upwards after gazing at the user instead of downwards.

We had to disable the break after an unsuccessful attention phase in the interactive version as the interaction duration would have been significantly shorter and thus not comparable to the two non-interactive versions.

Study

For the study, we recruited 16 subjects, solely women due to the male virtual counterpart, and presented them with all three interaction modes. The order of the three interaction modes (i.e. *interactive*, *non-interactive ideal* and *non-interactive anti-flirt*) was randomized for each subject to avoid any bias due to ordering effects. The procedure was as follows: First, the subjects were asked to fill in the first part of the questionnaire about demographic data. After placing the subjects in front of the eye tracker, a calibration, which took less than 2 minutes, was carried out. The subjects were told that they would be presented with a flirting agent and should try to engage in a flirt with the agent themselves using their gaze. They were informed that they would have to run the interaction sequence three times, but they did not know that there were different modes of interaction. After accomplishing one interaction sequence, the subjects had to fill in a post-sequence questionnaire about the interaction with the virtual agent. The study took about 20 minutes including the calibration for the eye tracker and answering the questionnaire.

Questionnaire

The post-sequence questionnaire (see Appendix C.4) used 13 attitude statements with a 5-point Likert scale to evaluate how the participants perceived the interaction with the system. The questions were related to the user's engagement of the interaction (five questions: E1 – E5), the exclusion of external influences (three questions: I1 – I3) and the quality of the gaze behavior model (five questions: Q1 – Q5).

Results

The analyses of the questionnaires were based on the one-way analysis of variance (ANOVA) across the different groups and the Tukey-HSD for the posthoc two-sided pairwise comparisons. Two subjects had to be excluded from the analysis due to technical difficulties with the eye tracker. In their case, the eye tracking data stream was discontinuous and thus the interactive version did not work properly.

Engagement of the Interaction The one-way ANOVA for the questions regarding the engagement revealed significant differences among the means for question E1 ($F(2, 39) = 3.02, p < 0.05, \eta^2 = 4.02$), E2 ($F(2, 39) = 6.07, p < 0.01, \eta^2 = 6.5$), E3 ($F(2, 39) = 5.98, p < 0.01, \eta^2 = 6.38$) and E5 ($F(2, 39) = 3.35, p < 0.05, \eta^2 = 2.31$). E4 did not reveal significant differences. The Tukey-HSD posthoc test for pairwise comparisons revealed a significant difference for E1 between the *interactive* and *anti-flirt* mode ($p < 0.05$), for E2 between the *interactive* and *anti-flirt* mode ($p < 0.01$) and for E3 between the *ideal* and *anti-flirt* mode ($p < 0.05$) and the *interactive* and *anti-flirt* mode ($p < 0.01$).

Questions E1 – E5 (see Figure 6.8) were related to the users' engagement in the gaze interaction with the virtual agent. All these questions resulted into a higher mean for the interactive mode, where the agent's gaze behavior was aligned to the user's gaze. In the interactive mode, the subjects rated Alfred's gaze behavior and mimics more realistic (E1) and enjoyed the interactions with Alfred more (E2). Furthermore, they uttered a higher interest in continuing the interaction with Alfred (E3) and in actually engaging in a conversation with him (E4). They also thought that their own interaction behavior was more natural in the interactive mode than in the two non-interactive modes.

Exclusion of External Influences The one-way ANOVA for the questions (I1 – I3) regarding the exclusion of external influences revealed no significant differences

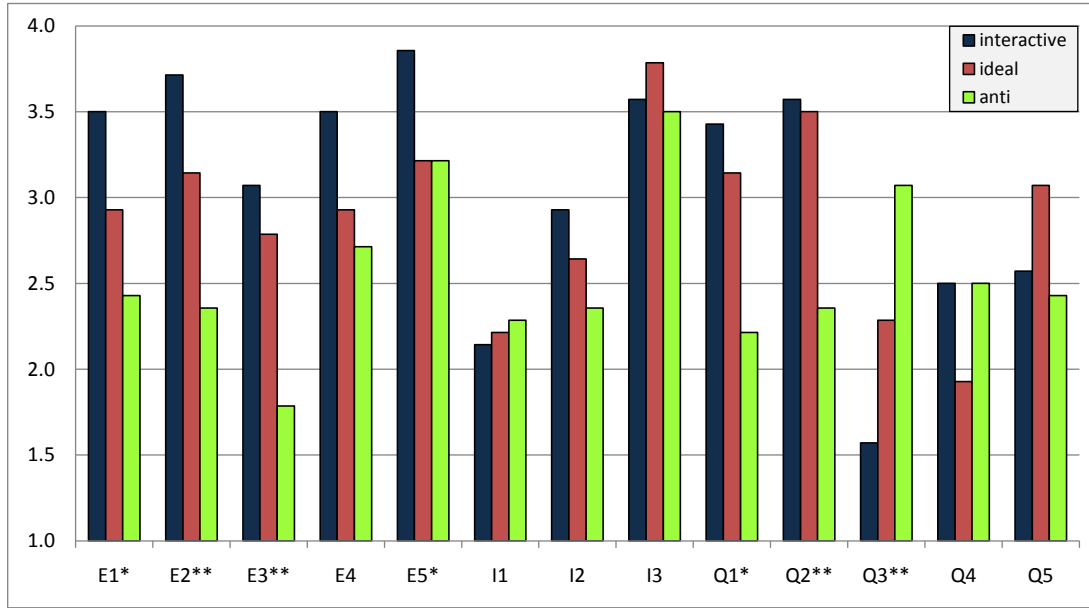


Figure 6.8: Results of the questionnaire for the interactive, non-interactive ideal and non-interactive anti-flirt mode (* $p < 0.05$, ** $p < 0.01$).

(see Figure 6.8). Unaffected by the mode of interaction, the subjects felt hardly watched by the equipment (I1). Furthermore, they gave similar subjective ratings for their flirting capabilities independent from the mode of interaction (I2) which we take as further evidence that they did not feel more disturbed in the interactive version. Finally, in all three modes of interaction, the subjects had the feeling that Alfred was looking into their eyes when he was looking at them (I3). Consequently, we can exclude artefacts due to different user sizes and the resulting different positions relative to the agent.

Quality of the Gaze Behavior Model The one-way ANOVA tests reveal significant differences for all questions regarding the quality of the gaze behavior model, for question Q1 ($F(2, 39) = 4.27, p < 0.05, \eta^2 = 5.64$), Q2 ($F(2, 39) = 5.27, p < 0.01, \eta^2 = 6.5$) and Q3 ($F(2, 39) = 6.00, p < 0.01, \eta^2 = 7.88$). The Tukey-HSD posthoc test for pairwise comparisons reveals significant difference for Q1 between the *interactive* and *anti-flirt* mode ($p < 0.05$), for Q2 between the *ideal* and *anti-flirt* mode ($p < 0.05$) and the *interactive* and *anti-flirt* mode ($p < 0.05$) and for Q3 between the *interactive* and *anti-flirt* mode ($p < 0.01$).

Questions Q1 and Q2 referred to the ability of the virtual agent to convey interest through its gaze behavior and mimics. The subjects had the impression that Alfred was more interested in them (Q1) and flirting more with them (Q2) in the interactive version than in the other two versions. Nevertheless, they did not have the feeling that Alfred's gaze behaviors were obtrusive (Q3).

Q4 and Q5 did not reveal any significant differences. The subjects felt that the agent gazed at them directly (Q4) and was interactive (Q5) with a medium score in all three conditions. Surprisingly, the user did not perceive the interactive agent as more interactive even though this agent was more positively rated (E1 – E5).

Analyzing the User's Gaze Behavior During the Interaction

In the following, we perform a more careful analysis of the interactive mode. Here, the overall interaction took between 59.1 and 107.8 seconds, 82.6 seconds on average. The first part of the interaction was completed on average after 24.5 seconds (17.1 seconds minimum and 32.0 seconds maximum) which corresponds to the observation by [Tramitz \(1992\)](#) that the first 30 seconds of an interaction lay the foundations for further interactions.

As it turned out, the user was more pro-active than the agent in establishing and breaking off eye contact. In 76.9 % of the cases, the gaze contacts were initiated by the human user. In 85.7 % of the cases, the human user decided to break off eye contact.

Four subjects did not execute downward gazes after breaking off eye contact which are typical of flirting situations. Also the remaining candidates showed this behavior only one to three times.

Not every attempt to establish gaze contact also led to mutual gaze. Each subject experienced at least once (at a maximum five times) the situation that the agent did not respond to his or her attempt to establish mutual gaze. As a reason we indicate that the user was averting his gaze immediately after meeting the agent's eyes. That is in some cases, Alfred's response came too late. On the other hand, in 12 out of 14 interactions, it happened only once that the agent tried in vain to establish eye contact.

Discussion

Overall, the experiment led to promising results. In the interactive and the ideal mode, the agent was able to show the users that he had an interest in them and the users also had the feeling that he was flirting with them (Hypothesis I). Furthermore, we found that the effect was increased when moving from the ideal to the interactive mode (Hypothesis II). Although significant differences were only detected between the interactive and the anti-flirt version, the means of the interactive version were always rated higher than the ideal version and the means of the ideal version were always rated higher than the anti-flirt version. In addition,

the experiment revealed that the interactive version contributed to the user's enjoyment, increased their interest to continue the interaction or even to engage in a conversation with Alfred even though the differences were only significant for the interactive and the anti-flirt version (Hypothesis III). The users did not have the feeling that the agent was significantly more responsive in the interactive than in the non-interactive versions. The result is in conflict with a result we obtained for an earlier experiment with an eye-gaze controlled agent. In the earlier experiment, the interactive agent was perceived as more responsive than the non-interactive agent. The subjects felt, however, also more disturbed by the perceptive agent (Eichner et al., 2007). Obviously, the users enjoyed the interactive version more and found it more engaging without perceiving it, however, as more interactive. The higher level of engagement was also reflected by the users' behavior. We were seeking more often for eye contact with the agent than in the non-interactive versions. Furthermore, the subjects found the interactive agent more realistic and indicated that it was more natural to interact with it.

6.4 Experiment II: Verbal Interaction

Implementing the interactive storytelling (IS) concept involves many computing technologies: virtual or mixed reality for creating the artificial world, and artificial intelligence techniques and formalisms for generating the narrative and characters in real time. As a character in the narrative, the user communicates with virtual characters much like an actor communicates with other actors. This requirement introduces a novel context for multimodal communication as well as several technical challenges. Acting involves attitudes and body gestures that are highly significant for both dramatic presentation and communication. At the same time, spoken communication is essential to realistic interactive narratives (Cavazza et al., 2009).

A large variety of interfaces have been proposed for interactive storytelling including desktop-based interfaces as well as novel forms of interaction based on the use of electronic toys, conversation with virtual characters or instrumented story environments. For example, the eCIRCUS project investigates natural language conversation with virtual characters in FearNot! (Aylett et al., 2006) as well as various forms of bodily and tangible interactions including interaction with a pressure sensitive dancing pad, gesture-based interaction with Nintendo's WiiMote and tangible interaction using mobile phones in ORIENT (Aylett et al., 2009). Apart from our earlier work (Cavazza et al., 2009) where we developed a story character that responds to the user's emotive tone, there is, however, hardly any con-

versational interface to interactive storytelling that emphasizes the socio-emotive aspects of interaction and integrates sophisticated technologies to recognize the user's emotive state.

The background narrative for this system is an adaptation of three chapters of the XIXth century classic *Madame Bovary* by Gustave Flaubert (Flaubert, 1856). Emma Bovary is married to a country doctor, Charles Bovary, but boredom in her married life has drawn her towards Rodolphe Boulanger. The user plays the role of Rodolphe who may address Emma or respond to her complaints and love declarations by using free-style spoken natural language input.

In the following, we describe how the user's speech and gaze behaviors is analyzed using a framework for the synchronized analysis of multimodal input. After that, we present two gaze models that are both informed by studies of human gaze behaviors: an *interactive* gaze model that is sensitive to the user's gaze and a *non-interactive* gaze model that does not have the information on where the user is looking. Next, we report on a study we conducted within the this interactive storytelling system in order to compare the two gaze models focusing on the users' experience and their attitude towards the agent. And finally, we analyze the users' gaze during the speech dialogs with the virtual character.

6.4.1 Analysis of Conversational and Social Behaviors

Unlike earlier systems (Dow et al., 2007), our focus is not on the analysis of the semantics, but on the socio-emotional aspects of such a conversation. To analyze the user's behaviors when interacting with the virtual character, we employed a framework for multimodal signal processing in real-time (Wagner et al., 2009) and extended it by dedicated algorithms for recording and analyzing the user's eye gaze.

Architecture

As depicted in Figure 6.9 the framework mediates between the sensors, which capture the user interaction, and the system, which generates in real-time the response according to the input. The information provided by our framework ranges from raw sensor data, such as eye coordinates or skin conductivity level, over low level features, such as voice pitch or heart rate, to high level description, such as the level of interest or emotional states. Exchange of information to the character control system happens continuously based on a regular update interval, or discrete, either driven by the signals, for example based on activity

detection, or on request, for example when a decision has to be made. For the work presented here, we do not make use of all channels the framework supports. Rather, we focus on the acoustic properties of speech and on the user's gaze behaviors.

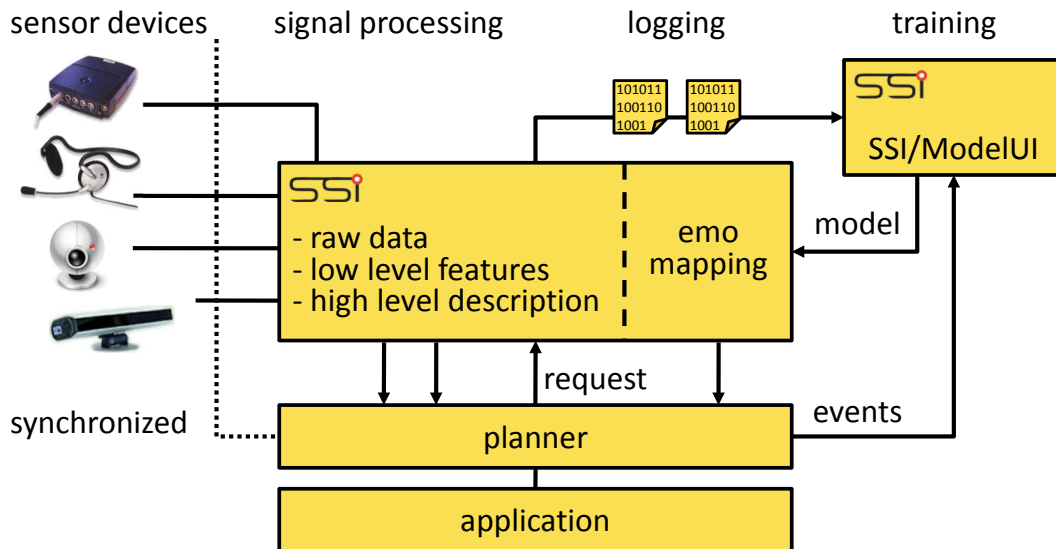


Figure 6.9: We measure user interaction with different sensor devices, which are synchronized and pre-processed through the framework.

Emotional categories extracted from the user's utterance are analyzed in terms of the current narrative context to produce a specific influence on the target character, which will become visible through a change in its behavior, achieving a high level of realism for the interaction. The character's behavior is driven by an emotional planner, which determines the actions a character may undertake based on its feelings. In addition to analyzing the acoustic of speech as input to the emotional planner, we track the user's gaze. So far, we do not make use of gaze to drive the narrative. Rather, we focus on gaze as a means to make users feel that the character is aware of them. That is the user's gaze has a direct impact on the character's behavior who would, for example, avert her gaze if the user continuously stares at her, see Section 6.2.

Emotional Speech

Affective input from the voice is analyzed by EmoVoice (Vogt et al., 2008), which has been integrated as a tool box into our framework. Real-time recognition of vocal emotions is a three-step process. First, the acoustic input signal coming continuously from the microphone is segmented into chunks by Voice Activity Detection (VAD), which segments the signal into speech frames with no pauses within

longer than about 0.5 seconds. Next, from this speech frame, a number of features relevant to affect are extracted. The features are based on pitch, energy, Mel Frequency Cepstral Coefficients (MFCC), the frequency spectrum, the harmonics-to-noise ratio, duration and pauses. The actual feature vector is then obtained by calculating statistics (mean, maximum, minimum, etc.) over the speech frame ending up with around 1300 features. A full account of the feature extraction strategy can be found in (Vogt et al., 2008).

In the last step, the feature vector is classified into an affective state. Integrated classifiers are Support Vector Machines (SVM) and Naïve Bayes (NB), while the latter one is used more often because it is faster and thus responds better to real-time demands. The NB classifier is very fast, even for high-dimensional feature vectors, and therefore especially suitable for real-time processing. However, it has slightly lower classification rates than the SVM classifier which is a very common algorithm used in offline emotion recognition. In combination with feature selection and thereby a reduction of the number of features to less than 100, SVM is also feasible in real-time.

Gaze

Many systems investigating interactive models of visual attention make use of head trackers (for example, (Nakano et al., 2003) or (Sidner et al., 2004)). They are able to roughly assess in which direction the user is looking, but do not have more detailed information on the user's gaze direction. In our work, we make use of the SMI iView X RED eye tracker.

To find fixations this thesis makes use of the I-DT algorithm described by Salvucci and Goldberg (2000). According to I-DT, a fixation is detected when the eye coordinates of a frame lie within the distribution $disp$. For each frame $disp$ is calculated with the following formula: $disp = (max_x - min_x) + (max_y - min_y)$ where min_x , max_x , min_y and max_y are the minimum and maximum coordinate values of all points inside the frame. If $disp$ is beyond a certain threshold the current frame is detected as the beginning of a fixation and then expanded by following points until the threshold is exceeded. This marks the end of a fixation. The samples in the final window are averaged to a single fixation point. For our purpose a minimum length of 120 ms and threshold of 15 pixels have been found to give reasonable results.

6.4.2 Virtual Character

Our virtual character is a full body 3D character that synchronizes speech, facial displays, head and eye movements to converse with the human user. For rendering the character and its animations the Horde3D GameEngine (Augsburg University, 2007) is used. This time, University of Teesside adopted their female virtual character named Emma (see Figure 6.10), which was enhanced to use the FACS to synthesize a huge set of different facial expressions. Emma was created at the University of Teesside and is part of their interactive storytelling framework (see (Cavazza et al., 2007)). Further details on how to create facial expressions with a virtual character can be found in Section 4.2. In Chapter 4.1.2 you will find details about the implementation of the lip synchronization.

Gaze Model

The gaze model (see Section 6.2) was modified for this evaluation. We modeled two different gaze modes for our agent. In the *interactive* mode, the character looks for about 2 s (between 1 and 3 s) at the user before she averts her gaze again for about 4 s (between 2 and 6 s). Whenever the user is looking at Emma, she tries to establish mutual gaze and to hold it for about 1 s (between 0.75 and 1.25 s). In the *non-interactive* mode, the agent's gaze model is parameterized in such a way that the agent seems to randomly look at the user or avert its gaze, and the virtual character gazes on average for a period of 1 s (0-2 s) in any state. For both modes, the duration of gaze to and away from the user is slightly adapted depending on whether the agent is talking or listening to account for the fact people look more at the interlocutor when listening than when talking, see (Argyle and Cook, 1976).

6.4.3 Evaluation of the Gaze Models

In the following, we present the results of a study we conducted using the interactive system as a test bed in order to find out how users perceive a character that reacts to their gaze. In particular, we wanted to know whether the integration of an gaze model had any impact on the user's perception of social presence (P), their level of rapport with the character (R), their engagement (E), the social attraction of the character (A) and the subjective perception of the story (S).

Experimental Setting

We prepared an experimental setting to compare the two gaze models introduced in Section 6.2 *interactive* and *non-interactive* while users are interacting with a virtual character.

The user is placed in front of a table on which the eye tracker was placed. The eye tracker with an incline of 23° was installed 80 cm above ground and 140 cm away from the projection surface. The user is seated 60 - 80 cm in front of the eye tracker. In total the user is about 2 m away from the virtual agent, which is within the *social space* according to Hall (1963). The projection surface sizes 120×90 cm, which displays the virtual agent in life-size (see Figure 6.10). To avoid that the user automatically stares at the virtual agent (which would happen if it was placed in the center of the visual display), we placed it on the left side. To offer an enriched scene where the user has the choice to look away from the virtual agent, Emma was placed in the dining room of her house, which includes chairs and tables (see Figure 6.7).



Figure 6.10: Set-up for the interaction with Emma.

The procedure was as follows: First, the subjects were placed in front of the projection screen. Then the eye tracker was calibrated, which took less than 2 minutes.¹ The subjects were first informed about the background of the story. Then, they

¹To measure user engagement, we also connected users with skin conductance and blood volume pressure sensors and recorded their upper body. These data have, however, not yet been analyzed.

were told that they would enter the story in the role of Rodolphe who finds Emma alone in the salon and should try to engage her in a conversation. To exclude any side effects resulting from dynamically evolving stories of varying quality, we decided to use fixed story lines for the experiment. Thus, for the experiment, just EmoEmma's gaze behavior was automated, but see (Cavazza et al., 2009) for an experiment with EmoEmma which included automated emotion recognition from speech. We do not consider fixed story lines as a major problem in this particular case since Emma's verbal utterances were carefully chosen so that the users could in general make sense of them. In addition, the scenario chosen - the user in the role of Rodolphe is expected to approach Emma to start an affair with her - left the user with enough space for interpretation. In the experiment, Emma produced 12 turns pausing briefly (5-10 s) after each of them to give the user a chance to respond. Emma started with 'Hello Rodolphe, I am so delighted!' and the user could for example answer with 'Hello Emma, I feel just the same way!'. The whole process for each subject took about 20 minutes including the introduction to the story sequence whereby one interaction sequence took about 3 minutes. The order of the two gaze models (i.e. *interactive* and *non-interactive*) was randomized for each subject to avoid any bias due to ordering effects. Overall, we recruited 19 subjects (2 females and 17 males) with a mean age 25.3 (SD = 3.1) for the experiment. All subjects were native speakers of German.

Social Presence, Engagement and Interactional Rapport

The objective of the study was to find out whether the different modes had any impact on the subjects' experience ratings. In particular, we used a post-questionnaire (see Appendix C.5) with a 9-point rating scale (from strongly disagree to strongly agree) to assess the subjects' sense of social presence (P), their level of rapport with the character (R), their engagement (E), the social attraction of the character (A) and the subjective perception of the story (S).

Measures *Social Presence (P)*. We assessed the subjects' sense of social presence using the items "I had the feeling that Emma was aware of me.", "I had the feeling of personal contact to Emma.", "Emma was impersonal." (reverse coded), and "Emma was reserved." (reverse coded).

Rapport with the Character (R). The level of rapport with the virtual character was measured using the items "I would have liked to continue the interaction with Emma.", "Emma's behavior was natural.", "I had the feeling that Emma reacted on me.", and "Emma's behavior was synchronous to mine.".

Engagement (E). We indexed the user's level of engagement with the following two items: "I enjoyed the first meeting with Emma." and "I found it easy to flirt with Emma."

Social Attraction of the Character (A). The users' social attraction of the character was measured using "I had the feeling, that Emma was interested in me." and "Emma was sympathetic."

Perception of the Story (S). The subjective perception of the story was assessed using the items "I would like to know how the episode with Emma continues.", "I had no problems to empathize with the part of Rodolphe.", and "I had the feeling to influence the story with my gaze."

Results The significance analyses between the interactive gaze mode and the non-interactive mode were conducted using a paired two-tailed *t*-test. A look at Figure 6.11 reveals that all groups received more positive ratings for the *interactive* gaze model than for the *non-interactive* gaze model.

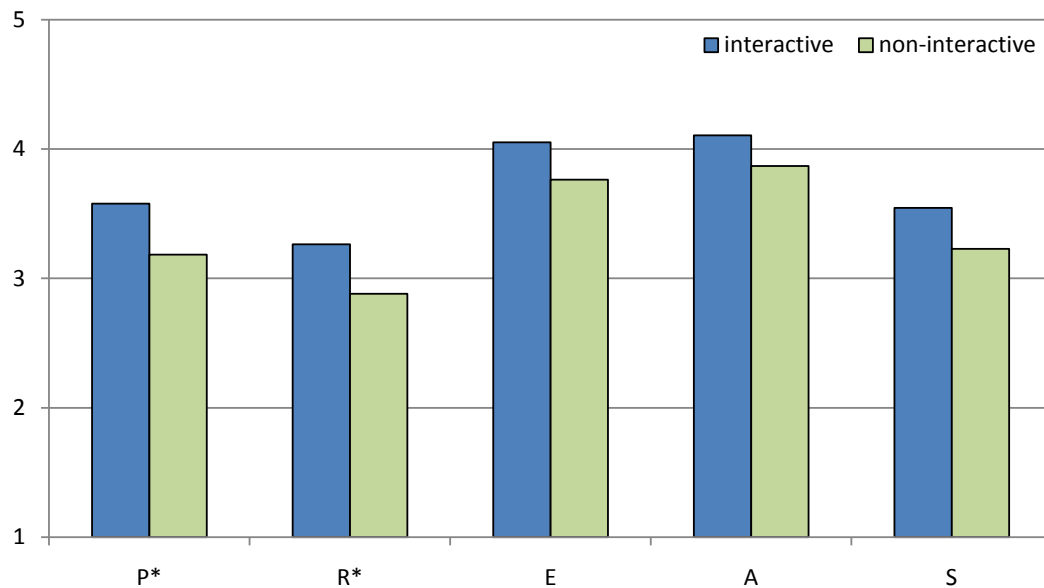


Figure 6.11: Results for the questions compared with the *interactive* and *non-interactive* gaze model while interacting with Emma (* $p < 0.05$).

The significance test reveals that the presence measure differs significantly between the *interactive* and *non-interactive* gaze mode (P: $t(75) = 2.6, p = 0.01, r = 0.29$). Also the rapport measure reveals significant differences between these two modes (R: $t(75) = 2.3, p = 0.02, r = 0.26$). However, the other measures did not reveal any significant differences (E: $t(37) = 1.6, p = 0.11$; A: $t(37) = 1.2, p = 0.25$; S: $t(56) = 1.5, p = 0.15$).

Analysis of the Subjects' Gaze Behaviors

First of all, we investigated to what extent the subjects were looking at Emma while she was speaking or silent. This gives us evidence whether the user interacts with Emma in a similar way as they would do in human-human interaction. We calculated the fixation points from the raw gaze data using the algorithm presented in Section 6.4.1. Furthermore, we divided the scene into two areas. The first area covers the eyes of the virtual character and the second area the rest of the scene.

We found that independent from the gaze mode, the users were looking at Emma around 76 % of the time in contrast to Kendon (1967) who found that in human-human interaction a human is looking on average 50 % of the time at an interlocutor. Further, Kendon reports that this quote varies from 28 % to 70 % whereas we found a variation of 46 % to 98 %.

(Argyle and Cook, 1976) found that humans look about 75 % at interlocutors while listening and 41 % while speaking. Independent from the gaze mode, we found that users interacting with a virtual agent look about 81 % of the time at the agent while listening and about 71 % of the time at Emma while speaking. Although the users were in total much more looking at Emma, the relationship between listening and speaking remains comparable (i.e. the user looks at the interlocutor considerably longer when listening than when speaking) to human-human interaction. These findings are in line with an study conducted by Rehm and André (2005) and they ascribe them to the novelty effect of the agent.

Considering a multimodal gaze model that takes the user's gaze and speech into account, we analyze where the users are looking when they start and stop speaking. We expect findings that can be integrated in a multimodal interactive gaze model for a virtual character that enables the agent to detect whether a user plans to say or answers something or is expecting further advices from the system. In this way an attentive system could recognize whether the current stimulus already suffices to expect an answer or feedback from the user or if the system needs to elaborate the current dialog part.

Figure 6.12 shows the gaze pattern when the users start speaking. We chose to analyze an 3.5 seconds interval, where we looked at the 3 seconds before the users started to speak and 0.5 seconds after the users started to speak. The users started to speak at $t = 0$ and we collected overall 430 utterances for this analysis. In Figure 6.12, three phases are shown: *Emma speaks*, *pause* and the *user starts speaking*. The pause after Emma speaks and the user answers is on average 1.43 seconds (SD = 1.05). The vertical axis indicates the users' current gaze target, where 0 means the user looks away and 1 that the user looks at Emma's face. On average, the users

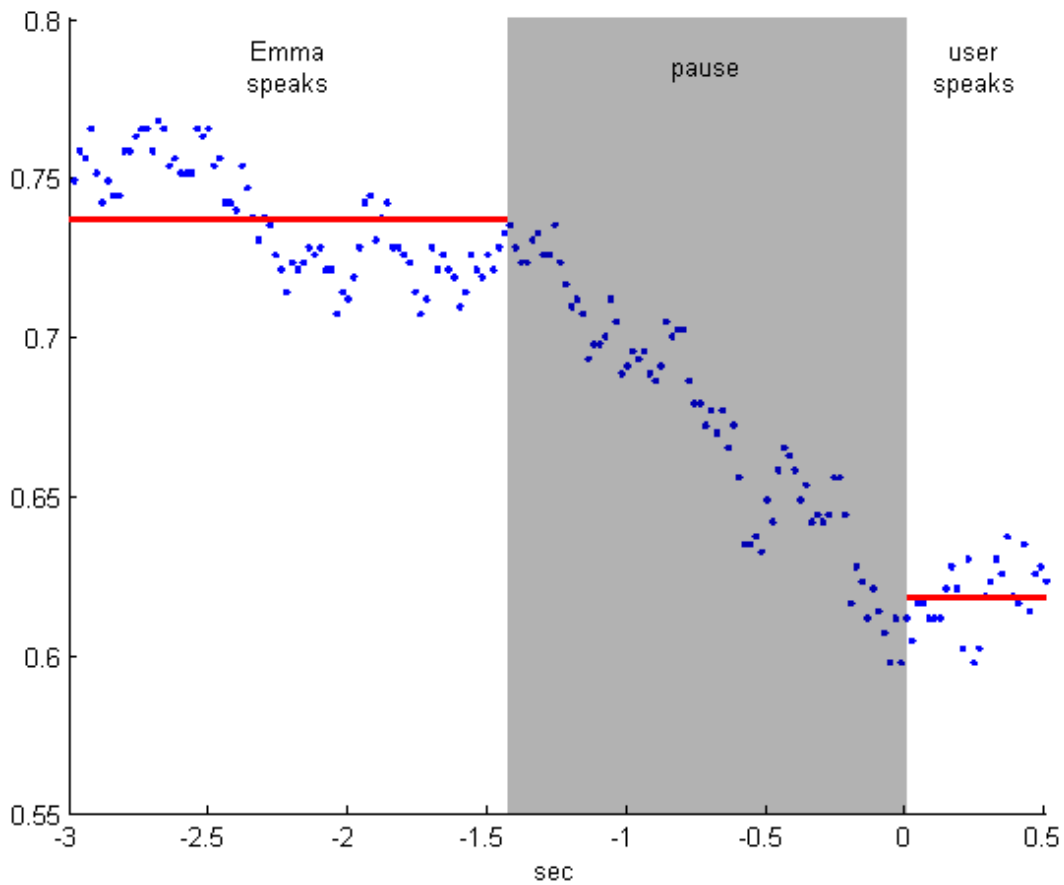


Figure 6.12: Gaze pattern before the users start speaking. The vertical axis indicates the gaze target (0 = looking away, 1 = looking at Emma, red line = average) during conversation. The user starts speaking at $t = 0$.

looked significantly more at Emma while she was speaking than when the users started to answer, where they averted their gaze ($t(102) = 32.8, p = 0, r = 0.96$). The finding is not only statistically significant, but has also a large effect (r) and so indicates a substantive finding. Morency et al. (2006) also found that users avert their gaze while thinking or answering.

In Figure 6.13 we plot the users' gaze pattern at the end of their utterance. We analyzed a 2.5 seconds interval, where we looked at 0.5 seconds before the users stop speaking and 2 seconds afterwards. The users stopped speaking at $t = 0$ and we collected overall 378 utterances from the users. The users started to look significantly more often at Emma face after they stopped speaking ($t(123) = 6.2, p = 0, r = 0.49$). Looking at the gaze pattern in Figure 6.13 reveals that after the users end their utterance, their gaze behavior looks like an increasing sawtooth pattern. Which means that they are rhythmically alternating their gaze between Emma's face and the rest of the virtual scene.

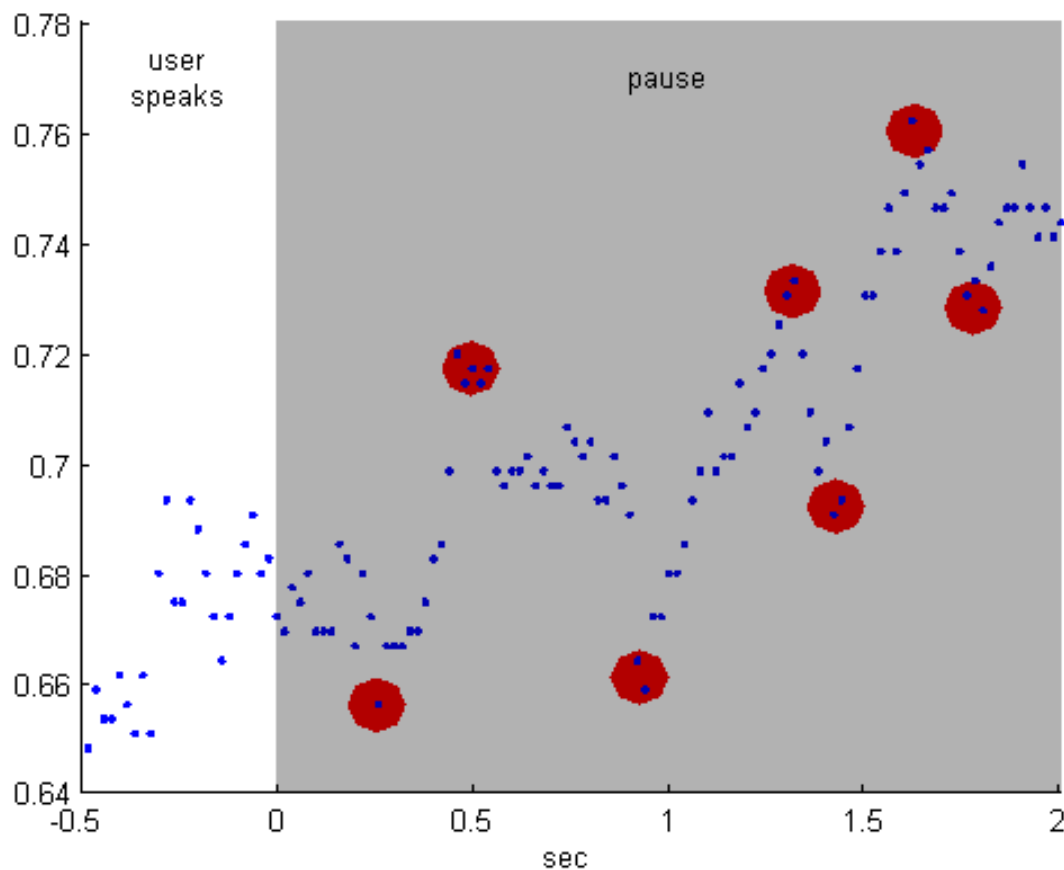


Figure 6.13: Gaze pattern after the users stopped speaking. The vertical axis indicates the gaze target (0 = looking away, 1 = looking at Emma). The user stops speaking at $t = 0$.

6.5 Conclusion

In Section 6.3, we presented an eye-gaze based interaction model for embodied conversational agents which incorporates studies of flirting in human-human interaction. The approach was tested using a 3D character that enables realistic gaze behaviors in combination with expressive mimics. We successfully integrated an eye tracker in a life-size application with a quite huge interaction screen display. The contact-free eye tracker has proven appropriate and reliable for a life-size interactive setting with a large interaction screen display. It was not perceived as intrusive or disturbing and thus the users could interact in a relatively natural manner. As we did not give the subjects special instructions how they should behave in front of the eye tracker, they moved freely without taking care of the eye tracker. The usage of such eye trackers is promising, as only 2 of 16 subjects had to be removed from the recordings as the eye tracker did not work trouble-free with them.

To enable smooth interactions, a high amount of alignment and coordination was required. In particular, the agent had to sense and respond to the user's gaze behavior in real-time. Despite the technical challenges involved in this task, the interactive version was perceived as more natural than the non-interactive versions. The user also had the impression that the agent had an interest in them without perceiving as obtrusive. Both subjective user ratings as well as objective user observations revealed that the users were more eager to continue interaction with the agent.

Usually, attractiveness is considered as a prerequisite for successful flirting. Our subjects rated Alfred as sympathetic, but also little attractive. Nevertheless, the incorporation of flirting tactics has proven beneficial. Thus, this can be taken as evidence that the flirting tactics as implemented in this work are of benefit to a much broader range of situations with agents than just dating, e.g. initiate human-agent interaction or regulating turn-taking in dialogues.

In Section 6.4, the gaze model was integrated and tested within an existing story telling system in which the user could freely interact with the main character impersonating one of the story characters. An evaluation provided interesting results regarding the users' perception of the interaction and their attitude towards the character. We found that the interactive gaze mode led to a better user experience compared to the non-interactive gaze mode. Indeed, the interactive gaze mode achieved a higher score for all items of a questionnaire measuring the user's sense of social presence, their level of rapport with the agent, their engagement, the social attraction of the character and the subjective perception of the story. These results are in line with our previous work (Bee et al., 2009a) where we analyzed the interaction with a virtual character based on gaze only. To improve the recognition and mapping of the spoken words and the user's current attention, the system could benefit from the integration of a method presented in (Prasov and Chai, 2008) or (Pfeiffer et al., 2009). Additionally, we found that users adhere to patterns of gaze behaviors for speaker and addressee that are also characteristic of dyadic human-human interactions. However, they looked significantly more often to the virtual interlocutor than is typical of human-human interactions.

Chapter 7

Conclusion

The work presented in this thesis focused on how to build gaze-based interfaces with emphasis on anthropomorphic interfaces. This chapter will give a short summary, list the main contributions of this thesis and finally will give an outlook on possible future work.

7.1 Summary

Gaze Aware Graphical User Interfaces The main aspect in creating gaze-based interactive interfaces is to distinguish between explicit and implicit interaction. For explicit interfaces it is important to take the nature of the eyes into account, which ends up in interfaces that lead the eye through it. Guidance and well structured information flow on a gaze-activated display plays an important role. For implicit gaze-activated interfaces it is important to understand the users' gaze pattern and to deliver the interpretations at the right time. It is inevitable to create a robust real-time system, that is capable to detect the user's gaze, interpret and react just in time.

Explicit gaze interaction has the potential of becoming a new form of interaction in human-computer interfaces. We developed a new writing system for gaze controlled interaction. Our very first prototype can easily compete with gaze-controlled keyboard-based systems. Based on the results of a user study, we formulated some guidelines for the design of such systems. The principles, (1) always move and (2) never lift the stylus, perfectly match the nature of human's gaze and should be considered in future designs for gaze interaction.

The implicit gaze-based system was able to detect subjects' choices correctly in 81 % of the cases. We compared recognition rates and decision times for 'different'

vs. 'similar' tie pairs. Recognition rates were 75 % (different ties) and 81 % (similar ties); decision times were 6.8 s (different ties) and 7.65 s (similar ties).

Perception of Facial Display The gaze direction plays an import role in showing affect and attention. A human's *emotional state* or *personality* can be strongly influenced by the *gaze* direction. This fact needs to be taken into account for creating gaze behavior for virtual humans. There is a strong effect on the virtual agent's gaze direction when combined with different facial expressions. The strength of virtual agent's dominance, for example, varied with the direction of the gaze. As well as the perception of emotion is influenced by the direction of gaze, the personality is also influenced by the gaze direction. This leads to quite complex relationships between affect and personality and the influence of gaze.

Higher dominance values were found for facial expressions conveying joy, anger and disgust. The dominance rating for a neutral facial expression, however, was significantly lower than that for joy, anger or disgust. Sadness and fear were perceived significantly less dominant in our experiment than joy, anger, disgust and neutral. Joy was perceived as less dominant when the gaze was averted. In contrast, anger and fear led to an increase in dominance in combination with averted gaze. Further, we found that gaze aversion had no influence on dominance ratings in combination with faces showing sadness, disgust or a neutral expression. Significant differences between an upward and downward directed head orientation could only be found for a neutral state, anger and sadness. Here, a lowered head orientation reduced the perception of dominance. Finally, we could show that an upward head orientation in combination with direct gaze was rated as significantly more dominant than a downward oriented head with averted gaze direction for anger and disgust. To summarize these findings, it matters where a virtual agent directs its attention dependent on its current affective state, and such effects need to be taken into account when modeling attentive affective agents.

Further, we investigated how the gaze and head orientation determines the perception of extraversion, agreeableness and emotional stability. The obtained results shed light on the question whether certain visual cues could be associated with personality traits. With the experiment we found that for the virtual character the up-side head orientation is related to extraversion, center-down head orientation is related to agreeableness, and center-side head orientation is related to emotional stability. We also confirmed our hypothesis that head side orientation, i.e. if the character's head or gaze is oriented to the left or to the right, does not influence the perception of personality traits. From our results we could also observe that people take into consideration more than a limited number of visual

cues to infer personality. Personality is not only influenced by head orientation or gaze. Take an “up-side” head and combine it with a facial expression of sadness, and the perception of personality could be other than extraversion. In this sense, and because of the nature of the study, we consider it necessary to perform more experiments related to these visual cues as well as other cues as facial traits (the physical characteristics of the face), gender, or facial expressions, in order to obtain a generalizable model.

Interactive Gaze Model Finally, this thesis presents an interactive gaze model for a virtual human that takes the user’s current gaze into account. This gaze model is capable to recognize and establish mutual gaze and thus is able to avoid staring. The users considered this interactive gaze model as better compared to purely inferred gaze models. It was not only considered as better in a non-verbal setting. We could show that it is also perceived better by the user in a verbal setting, where the user has to concentrate on the dialog act.

The non-verbal approach was tested using a 3D character that enables realistic gaze behaviors in combination with expressive mimics. We successfully integrated an eye tracker in a life-size application with a quite huge interaction screen display. The contact-free eye tracker has proven appropriate and reliable for a life-size interactive setting with a large interaction screen display. It was not perceived as intrusive or disturbing and thus the users could interact in a relatively natural manner. To enable smooth interactions, a high amount of alignment and coordination was required. In particular, the agent had to sense and respond to the user’s gaze behavior in real-time. Despite of the technical challenges involved in this task, the interactive version was perceived as more natural than the non-interactive versions. The user also had the impression that the agent had an interest in them without perceiving it as obtrusive. Both subjective user ratings as well as objective user observations revealed that the users were more eager to continue interaction with the agent.

The gaze model for the verbal setting was integrated and tested within an existing story telling system in which the user could freely interact with the main character impersonating one of the story characters. An evaluation provided interesting results regarding the users’ perception of the interaction and their attitude towards the character. We found that the interactive gaze mode led to a better user experience compared to the non-interactive gaze mode. Indeed, the interactive gaze mode achieved a higher score for all items of a questionnaire measuring the user’s sense of social presence, their level of rapport with the agent, their engagement, the social attraction of the character and the subjective perception of the story. Additionally, we found that users adhere to patterns of gaze behaviors for speaker

and addressee that are also characteristic of dyadic human-human interactions. However, they looked significantly more often to the virtual interlocutor than is typical of human-human interactions.

7.2 Contributions

Parts of this work were presented on TV (DW TV)¹, watched more than 17.000 times on YouTube², mentioned in the local press³, won two times the IVA GALA award at the international conference on Intelligent Virtual Agents in 2008⁴ and 2009⁵, and was nominated for the best paper award in 2009. Most of the technical part of this thesis is available as source for the public and is also well documented in a public wiki⁶. Further, this work was not only part of a lot university courses and student theses but also contributed to several research projects.

7.2.1 Methodological Contributions

The presented methodologies can easily be replicated by other works. Hence, it should be no problem to e.g. reproduce the measuring of the users' perceived affective states by using the three dimensional PAD model or to reproduce the measuring of the perceived interactivity during agent-human interaction.

Study I: Understanding how to develop explicit and implicit gaze-controlled interfaces (see Section 3.1 and Section 3.2). One main aspect in creating gaze-based interactive interfaces is to distinguish between explicit and implicit interaction and to take the nature of the eye into account.

Study II: Understanding how to measure users' perceive affection and personality in anthropomorphic interfaces. The methods presented in Chapter 5 show how to validate variations in an anthropomorphic interface.

¹<http://www.dw-world.de/dw/article/0,,6506477,00.html>

²<http://youtu.be/aCS--pxeXT4>

³<http://www.augsburger-allgemeine.de/landsberg/Gute-Miene-zum-virtuellen-Spiel-id3975741.html>

⁴http://hmi.ewi.utwente.nl/gala/finalists_2008

⁵http://hmi.ewi.utwente.nl/gala/finalists_2009

⁶<http://hcm-lab.de/projects/GameEngine/>

Study III: Understanding how to validate an interactive anthropomorphic interface during non-verbal and verbal human-agent interaction. In Chapter 6 we presented a validation framework for testing the interactivity and user's perception of a human-agent framework which is not limited to gaze-based human-agent interaction.

7.2.2 Theoretical Contributions

Study I: Understanding how humans gather information through gaze and how this can be utilized to create gaze-based graphical user interfaces (see Section 3.1 and Section 3.2).

Study II: Understanding how head orientation and gaze transmits affective and personality traits (see Section 5.1 and Section 5.2). Evidence that head orientations influence the perception of affection and personality.

Study III: Understanding how gaze aware anthropomorphic interfaces need to be designed for non-verbal and verbal gaze interaction (see Section 6.3.1 and Section 6.4). Evidence for the necessity of an interactive gaze model in natural human-agent interaction.

7.2.3 Practical Contributions

Study I: Computational model for preference detection that lets a system detect the user's preference during explicit and implicit gaze interaction (see Section 3.1 and Section 3.2). A gaze interactive framework that lets create GUIs, detect the user's eye gaze with fixation detection and lets the user gaze interact with the GUI in explicit and implicit manner.

Study II: Parameters on how an anthropomorphic interface is able to control affect (see Section 5.1) and personality (see Section 5.2) through altering facial expressions, head and gaze orientation. Tools that simplified the creation of facial expressions of a virtual character.

Study III: Computational model and parameters on how to align agent-user based gaze interaction (see Section 6.3 and Section 6.4).

7.3 Future Work

Based on this thesis, future work should handle specific gaze related topics. Bee and André (2008a) could show that culture has a great impact on gaze behavior in human-human interaction. For example, staring can be considered as impolite in one culture and polite and necessary in another culture. These cultural differences need to be taken into account when creating gaze behavior models. It would be interesting to investigate how cultural differences in gaze behavior can be transferred to the interaction with virtual characters.

While this thesis and related work could show the state in dominance perception of different gaze directions, there is still a big gap in the research of personality perception and the role of gaze. It is still open how much and how gaze influences the perception of personality. In addition, while this thesis could only investigate the role of static head poses and gaze directions, it would be straightforward to apply the presented methodologies to animated head poses and to measure the user's perception of these. This is important to refine the model for human-agent interaction.

Bee et al. (2010a) could show that speech (what is said) and gaze behavior can alter the user's perception of dominance. While this work only demonstrated a first step, it there could be done much more in combination with emotions and personality.

Finally, this thesis introduced a systematic investigation of gaze based interaction with virtual humans. As human-like robots become more and more available, it would be straightforward to apply the presented methods for human-robot interaction and investigate how humans perceive specific robot's behavior. A first step was made in (Häring et al., 2011), which examined with the help of the PAD model the affective perception of a human-like robot's speech and gesture behavior.

Appendix A

Facial Expression XML

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

```
<Expression>
```

```
  <!-- Emotional Expressions -->
```

```
  <FacialExpression name="neutral">
```

```
    <AU id="15" value="0.2" />
```

```
  </FacialExpression>
```

```
  <FacialExpression name="joy">
```

```
    <AU id="6" value="1" />
```

```
    <AU id="12" value="0.8" />
```

```
    <AU id="20" value="0.2" />
```

```
    <AU id="27" value="0.1" />
```

```
  </FacialExpression>
```

```
  <FacialExpression name="sadness">
```

```
    <AU id="1" value="1" />
```

```
    <AU id="4" value="1" />
```

```
    <AU id="7" value="1" />
```

```
    <AU id="11" value="1" />
```

```
    <AU id="15" value="0.1" />
```

```
    <AU id="17" value="0.8" />
```

```
  </FacialExpression>
```

```

<FacialExpression name="fear">
  <AU id="1" value="0.5" />
  <AU id="2" value="1" />
  <AU id="4" value="0.5" />
  <AU id="5" value="0.8" />
  <AU id="20" value="0.5" />
</FacialExpression>

<FacialExpression name="surprise">
  <AU id="1" value="0.5" />
  <AU id="2" value="1" />
  <AU id="5" value="1" />
  <AU id="27" value="0.6" />
</FacialExpression>

<FacialExpression name="anger">
  <AU id="4" value="0.9" />
  <AU id="5" value="1" />
  <AU id="7" value="0.9" />
</FacialExpression>

<!-- Visemes -->

<FacialExpression name="ae_ax_ah_aa_ao_er_ay">
  <AU id="10" value="0.1520468" />
  <AU id="20" value="0.3988304" />
  <AU id="27" value="0.748538" />
</FacialExpression>

<FacialExpression name="ey_eh_uh">
  <AU id="18" value="0.3333333" />
  <AU id="27" value="0.6491229" />
</FacialExpression>

<FacialExpression name="aw_y_iy_ih_ix_h_k_g_ng">
  <AU id="10" value="0.1169591" />
  <AU id="25" value="1" />
  <AU id="27" value="0.3625731" />
</FacialExpression>

```

```
<FacialExpression name="w_uw">
  <AU id="18" value="0.4690059" />
  <AU id="27" value="0.4736842" />
</FacialExpression>

<FacialExpression name="ow">
  <AU id="18" value="0.1754386" />
  <AU id="27" value="0.5263158" />
</FacialExpression>

<FacialExpression name="oy">
  <AU id="10" value="0.2" />
  <AU id="18" value="0.2163743" />
  <AU id="27" value="0.5614035" />
</FacialExpression>

<FacialExpression name="r">
  <AU id="10" value="0.2" />
  <AU id="20" value="0.04093567" />
  <AU id="25" value="0.8304093" />
  <AU id="27" value="0.1871345" />
</FacialExpression>

<FacialExpression name="l">
  <AU id="10" value="0.2" />
  <AU id="20" value="0.1766082" />
  <AU id="25" value="0.4912281" />
  <AU id="27" value="0.1871345" />
</FacialExpression>

<FacialExpression name="s_z_t_d_n">
  <AU id="10" value="0.2" />
  <AU id="12" value="0.2222222" />
  <AU id="20" value="0.2690058" />
  <AU id="25" value="1" />
</FacialExpression>

<FacialExpression name="sh_ch_jh_zh">
```

```
<AU id="10" value="0.2" />
<AU id="16" value="0.2690058" />
<AU id="18" value="0.4853801" />
<AU id="25" value="1" />
</FacialExpression>

<FacialExpression name="th_dh">
  <AU id="10" value="0.2" />
  <AU id="16" value="0.04327485" />
  <AU id="25" value="0.8128655" />
  <AU id="27" value="0.2163743" />
</FacialExpression>

<FacialExpression name="f_v">
  <AU id="10" value="0.3" />
  <AU id="23" value="0.128655" />
  <AU id="25" value="0.3508772" />
</FacialExpression>

<FacialExpression name="p_b_m">
  <AU id="24" value="0.4502924" />
</FacialExpression>

</Expression>
```

Appendix B

Horde3D GameEngine Configuration for Alfred

```
<!DOCTYPE HordeSceneGraph>
<Group name="Alfred" >
  <Camera pipeline="pipelines\forward.pipeline.xml"
    topPlane="0.0254897" bottomPlane="-0.0254897"
    tx="0" ty="4" tz="15"
    sx="1" sy="1" sz="1"
    ry="0" rx="0" rz="0"
    leftPlane="-0.0339862" farPlane="1000" rightPlane
      ="0.0339862" nearPlane="0.1"
    name="Camera" >
    <Attachment type="GameEngine" name="Camera" >
      <Socket protocol="UDP" port="5553" address="
        127.0.0.1" type="server" />
      <SoundListener/>
    </Attachment>
  </Camera>

  <Light lightingContext="LIGHTING"
    tx="0" ty="10" tz="15"
    rx="-25" ry="0" rz="0"
    col_B="0.96" col_G="0.96" col_R="0.96"
    name="light1" material="materials/light.material.
      xml" shadowContext="SHADOWMAP" shadowMapBias="
        0.15" shadowMapCount="0" radius="25" fov="90"
    shadowSplitLambda="0" />
```

```

<Reference
    tx="0"      ty="-2.3"  tz="-1.6"
    sx="0.15"   sy="0.15"  sz="0.01"
    rx="0"      ry="0"     rz="0"
    sceneGraph="models\background\hedge.scene.xml" name
        ="background" />

```

```

<Reference
    tx="0"  ty="0"  tz="0"
    rx="0"  ry="180" rz="0"
    sx="1"  sy="1"  sz="1"
    sceneGraph="models\alfred\alfred_model.scene.xml"
        name="Alfred" >
    <Attachment type="GameEngine" name="Alfred" >
        <IK reye="Bone_Auge_R" leye="Bone_Auge_L"
            file="models\alfred\alfred_model.scene.xml"
            " neck="Bone_Head" />
        <MorphtargetAnimation/>
        <FACSControl file="models\expressions.xml" />
        <IdleBehavior/>
        <Socket protocol="UDP" port="5554" address="
            127.0.0.1" type="server" />
        <Sound3D phonemes="test.phonemes.xml" loop="0"
            " visemefile="models\visemes.xml" file="
            test.wav" gain="1" />
    </Attachment>
</Reference>

```

```

<Reference
    sx="1.05" sy="1"  sz="1.05"
    tx="0"    ty="0"  tz="0"
    rx="0"    ry="180" rz="0"
    sceneGraph="models\alfred\anzug.scene.xml" name="
        Suit" >
    <Attachment type="GameEngine" name="Suit" />
</Reference>

```

```

<!-- INVISIBLE PICKING OBJECTS -->

```

```

<!-- EYES -->
<Reference
  sx="0.5"    sy="0.5"    sz="0.5"
  tx="-0.025" ty="1.44"  tz="0.06"
  rx="0"      ry="0"      rz="0"
  sceneGraph="models\sphere\sphere.scene.xml" name=
    "sphereEyeL" >
  <Attachment type="GameEngine" name="sphereEyeL" >
    <LinkedObject name="Eye_L" />
  </Attachment>
</Reference>
<Reference
  sx="0.5"    sy="0.5"    sz="0.5"
  tx="0.025"  ty="1.44"  tz="0.06"
  rx="0"      ry="0"      rz="0"
  sceneGraph="models\sphere\sphere.scene.xml" name=
    "sphereEyeR" >
  <Attachment type="GameEngine" name="sphereEyeR" >
    <LinkedObject name="Eye_R" />
  </Attachment>
</Reference>
</Group>

```


Appendix C

Questionnaires

C.1 Facial Expression Control

Allgemeine Fragen

Alter: ____ Jahre

Geschlecht (bitte ankreuzen):

☐ männlich

☐ weiblich

Beruf:

☐ Student, Studienfach: _____

☐ anderer, nämlich: _____

	keine	wenig	mittel	ziemlich viel	sehr viel
Erfahrung mit Gamepads (Spielkonsolen, etc.):	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Erfahrung mit 3D Modellierung:	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Erfahrung mit Gesichtsanimation:	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Erfahrung mit Emotionsforschung:	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Fragebogen

zum System: Slider / Gamepad / Datenhandschuh

	stimmt nicht	stimmt wenig	stimmt mittel	stimmt ziemlich	stimmt sehr
1. Es macht Spaß das System zu benutzen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Ich musste oft zwischen der GUI und Alfred hin- und herblicken.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Ich war oft überrascht, wie das System auf meine Eingabe reagierte.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Es war körperlich anstrengend das System zu bedienen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Die Steuerung war für mich nachvollziehbar.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Das System bot mir genügend Einstellungsmöglichkeiten, um meine Vorstellungen umzusetzen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Die Genauigkeit der Einstellungen war für mich ausreichend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Ich musste viel probieren, bis ich die passenden Einstellungsmöglichkeiten finden konnte.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

C.2 Perception of Dominance

Willkommen - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

http://alfred.qtype.de/

Deutsch

Multimedia Concepts and Applications
Institute of Computer science

Erstmal Hallo und Danke für's Vorbeischauen!

Der Fragebogen besteht aus zwei Teilen, einem statistischen (hier rechts) und einer Reihe von Bildern die Du beurteilen sollst. Dafür stehen Dir pro Bild sieben Wortpaare zur Verfügung. Die Wortpaare klingen evtl. manchmal etwas komisch, aber jeweils eins der Wörter ist für das Bild sicher zutreffender als das andere. Und je zutreffender ein Wort ist, desto näher setzt Du das Knöpfchen auf der Skala darunter auf die jeweilige Seite. Hier ein Beispiel:

warm ... kalt

○ ○ ○ ○ ○ ○ ○ ○

Schön wäre es, wenn Du mir mindestens 10 Bilder beurteilen könntest. Danach taucht unter dem Formular folgendes Knöpfchen auf, das hoffentlich selbsterklärend ist:

Genuß - es reicht!

Wenn Du also keine Lust mehr hast, klick einfach drauf (Natürlich OHNE die Fragen zu dem aktuellen Bild ausgefüllt zu haben, denn die werden dann NICHT mehr gespeichert).

Danach kannst du auf der letzten Seite noch andere Leute einladen, die evtl. auch ein paar Minuten opfern würden um den Fragebogen auszufüllen.

Noch ein Hinweis, weil einige Leute Probleme mit dem Fragebogen hatten: Auf jedem Bild siehst Du einen virtuellen Charakter (namens Alfred) und die Wortpaare beschreiben immer das Verhältnis zwischen ihm und einer / seiner abstrakten Umgebung.

Statistische Daten

Alter: 28

Geschlecht: männlich

Heimatland: Deutschland

☒ Student (Fach): Informatik

☐ Beruf:

Erfahrung mit 3D Modellierung: viel

Erfahrung mit Gesichtsanimation: wenig

Erfahrung mit Emotionsforschung: mittel

Weiter >>>

Bilder bewerten - Mozilla Firefox

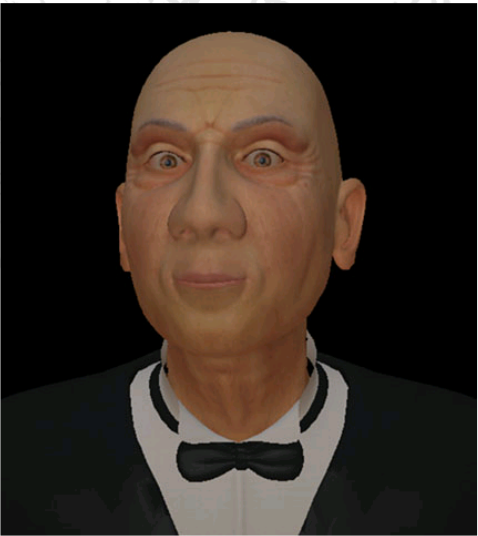
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

http://alfred.qtype.de/bilder.php?sessionID=818edf0ee1a8959fc1486136188f5da&spracheID=de

Deutsch

Multimedia Concepts and Applications
Institute of Computer science

Bild 1 / 10



kontrolliert ... kontrollierend

○ ○ ○ ○ ○ ○ ○ ○

einflussreich ... ehrfurchtsvoll

○ ○ ○ ○ ○ ○ ○ ○

beeinflusst ... beeinflussend

○ ○ ○ ○ ○ ○ ○ ○

leitend ... umsorgt

○ ○ ○ ○ ○ ○ ○ ○

unterwürfig ... dominant

○ ○ ○ ○ ○ ○ ○ ○

selbstständig ... gesteuert

○ ○ ○ ○ ○ ○ ○ ○




Weiter >>>

W3C HTML 4.01

C.3 Perception of Personality

Questionnaire

You will find a number of personality traits that may or may not apply to the displayed virtual character. Please chose your answer for each statement to indicate the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to the displayed virtual chracter, even if one characteristic applies more strongly than the other.

Next >>

Questionnaire


0% 100%

English ▼

General


Age

Only numbers may be entered in this field




Gender

☒ Female ☐ Male



Nationality




<< Previous Next >>

Questionnaire

0% 100%

English ▼

Picture
Please look at this picture and answer the following questions.



Please rate the virtual character in relation to the following questions:

	Disagree strongly		Neither agree nor disagree			Agree strongly	
Extraverted, enthusiastic	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Critical, quarrelsome	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Anxious, easily upset	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Reserved, quiet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sympathetic, warm	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Calm, emotionally stable	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

?

<< Previous Next >>

C.4 Non-Verbal Interaction

Fragebogen:

Teilnehmer Nr.

Alter:

☐ Unter 20 ☐ 20 – 25 ☐ 25 – 30 ☐ 30 – 35 ☐ 35 – 40 ☐ 45 und älter

Beruf / Studienrichtung:

Familienstand:

☐ Single ☐ in fester Partnerschaft

Erfahrung in Interaktion mit virtuellen Charakteren

☐ keine ☐ wenig ☐ mäßig bis viel

Wie viele Stunden pro Woche verbringst Du mit PC- oder Konsole-Spielen?

☐ 0 ☐ 1 – 3 ☐ 4 – 7 ☐ 8 – 15 ☐ 16 und mehr

Flirten:

Kennst Du typische Flirtsignale? Woran erkennst Du, dass jemand an Dir interessiert ist?

.....
.....
.....
.....

Wer ergreift die Initiative beim Flirten?

☐ meistens ich ☐ oft ich ☐ ausgewogen ☐ eher der Mann ☐ fast immer der Mann

Wie schnell, glaubst Du, merkst Du normalerweise, dass sich jemand für Dich interessiert

☐ sehr schnell ☐ schnell ☐ bald ☐ eher spät ☐ sehr spät

Wie gut, glaubst Du, kannst Du flirten?

☐ sehr gut ☐ gut ☐ mittelmäßig ☐ eher schlecht ☐ ziemlich schlecht

Fragen zur Interaktion

Wie stark stimmst Du den folgenden Aussagen zu?

1: starke Ablehnung, 2: eher Ablehnung, 3: neutral, 4: eher Zustimmung, 5: starke Zustimmung

- | | 1 | 2 | 3 | 4 | 5 |
|---|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 1. Ich finde Alfreds Blickverhalten und Mimik realistisch. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 2. Ich hatte Spaß an den Interaktionen mit Alfred. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 3. Ich hätte gern noch länger mit ihm interagiert. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 4. Ich hätte mich gern noch mit Alfred unterhalten. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 5. Ich glaube, ich habe mich in der Interaktion natürlich verhalten. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 6. Ich fühlte mich beobachtet (durch Eye-Tracker, Kamera, anwesende Personen) | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 7. Mein Flirtverhalten in den Interaktionen war gut. / Ich glaube ich habe gut im Flirt-Test abgeschnitten. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 8. Ich hatte den Eindruck, Alfred ist an mir interessiert. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 9. Ich hatte den Eindruck Alfred flirtet mich an. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 10. Ich hatte den Eindruck, Alfred sieht mir direkt in die Augen, wenn er in meine Richtung blickt. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 11. Ich fühlte mich von Alfred (unangenehm) angestarrt. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 12. Alfred hat mich kaum direkt angesehen. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 13. Ich hatte den Eindruck, Alfred reagiert auf mich/ mein Blickverhalten. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Sonstige Anmerkungen:

.....

.....

.....

.....

Abschließende Einschätzung**Fragen zum virtuellen Charakter:**

Wie stark stimmst Du den folgenden Aussagen zu?

1: starke Ablehnung, 2: eher Ablehnung, 3: neutral,
4: eher Zustimmung, 5: starke Zustimmung

		1	2	3	4	5
1.	Ich finde Alfred attraktiv.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.	Ich finde Alfred interessant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.	Ich finde Alfred sympathisch.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Interview:

Hat Dich etwas am Verhalten von Alfred irritiert? Was?

.....

.....

.....

.....

.....

.....

Sonstige Anmerkungen:

.....

.....

.....

.....

.....

.....

C.5 Verbal Interaction

Spielanweisung:

Die Geschichte spielt Mitten im 19. Jahrhundert in Yonville einer kleinen verschlafenen Stadt in der Normandie in Frankreich. Emma Bovary ist mit dem Landarzt Charles Bovary verheiratet. Sie hatte sich mit dieser Heirat ein aufregendes gesellschaftliches Leben erhofft. Leider fand sie sich aber recht schnell in einem langweiligen Dorfalltag mit ihrem eher einfach lebenden Mann wider.

Bis Rodolphe, ein Grundbesitzer, in ihr Leben trat, der durch seine charmantes Auftreten ihre Aufmerksamkeit gewann. Rodolphe interessiert sich sehr für Emma und möchte sie für sich gewinnen, um mit ihr in die aufregende Stadt Paris zu flüchten.

Nun trittst du in das Spiel ein und übernimmst die Rolle von Rodolphe. Deine Aufgabe ist es, mit Emma zu flirten und sie so für dich zu gewinnen. Dabei bleibt dir einzig dein Blickverhalten, um Emma von dich zu gewinnen.

Ablauf:

Du wirst zuerst auf den Eye Tracker kalibriert. Danach hast du drei Versuche um mit Emma zu flirten.

Viel Erfolg!

Persönliche Angaben: _____

Alter: _____

Beruf/Studienrichtung: _____

Wie oft hast du schon mit virtuellen Charakteren interagiert?

☐ noch nie ☐ 1 mal ☐ 2 – 5 mal ☐ > 5

Wie viele Stunden pro Woche verbringst du ungefähr mit PC- oder Konsole-Spielen?

☐ 0 ☐ 1 – 2 ☐ 3 – 5 ☐ 6 – 15 ☐ > 15

Wie stark stimmst du folgenden Aussagen zu?
(0 = starke Ablehnung ... 4 = neutral ... 8 = starke Zustimmung)

[illegible]

Bibliography

- Adams, R. B. and Kleck, R. E. (2005). Effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion (Washington, D.C.)*, 5(1):3–11.
- Albrecht, I., Schröder, M., Haber, J., and Seidel, H. P. (2005). Mixed feelings: expression of non-basic emotions in a muscle-based talking head. *Virtual Reality*, 8(4):201–212.
- Arellano, D., Bee, N., Janowski, K., André, E., Varona, J., and Perales, F. J. (2011). Influence of head orientation in perception of personality traits. In *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, pages 1093–1094.
- Argyle and Cook (1976). *Gaze & Mutual Gaze*. Cambridge University Press.
- Argyle, M. and Dean, J. (1965). Eye-contact, distance and affiliation. *Sociometry*, 28:289–304.
- Argyle, M. and Ingham, R. (1972). Gaze, mutual gaze and proximity. *Semiotica*, 6(1):32–49.
- Argyle, M., Lefebvre, L., and Cook, M. (1974). The meaning of five patterns of gaze. *European Journal of Social Psychology*, 4(2):125–136.
- Arya, A., Jefferies, L. N., Enns, J. T., and DiPaola, S. (2006). Facial actions as visual cues for personality. *Journal of Visualization and Computer Animation*, 17(3-4):371–382.
- Ashmore, M., Duchowski, A. T., and Shoemaker, G. (2005). Efficient eye pointing with a fisheye lens. In *GI '05: Proceedings of Graphics Interface 2005*, pages 203–210, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada. Canadian Human-Computer Communications Society.
- Augsburg University (2007). Horde3D GameEngine. <http://hcm-lab.de/projects/GameEngine/>.

- Aylett, R., Louchart, S., Dias, J., Paiva, A., Vala, M., Woods, S., and Hall, L. (2006). Unscripted narrative for affectively driven characters. *IEEE Comput. Graph. Appl.*, 26(3):42–52.
- Aylett, R., Vannini, N., André, E., Paiva, A., Enz, S., and Hall, L. (2009). But that was in another country: agents and intercultural empathy. In *AAMAS '09: Proc. of The 8th International Conference on Autonomous Agents and Multiagent Systems*, pages 329–336, Richland, SC.
- Balci, K. (2004). Xface: MPEG-4 based open source toolkit for 3D facial animation. In *Proceedings of the working conference on Advanced visual interfaces, AVI '04*, pages 399–402, New York, NY, USA. ACM.
- Bee, N. and André, E. (2008a). Cultural gaze behavior to improve the appearance of virtual agents. In *IUI Workshop on Enculturating Interfaces (ECI)*.
- Bee, N. and André, E. (2008b). Writing with Your Eye: A Dwell Time Free Writing System Adapted to the Nature of Human Eye Gaze. In *Perception in Multimodal Dialogue Systems*, pages 111–122.
- Bee, N., André, E., and Tober, S. (2009a). Breaking the Ice in Human-Agent Communication: Eye-Gaze Based Initiation of Contact with an Embodied Conversational Agent. In *9th International Conference on Intelligent Virtual Agents (IVA)*, pages 229–242. Springer.
- Bee, N., Falk, B., and André, E. (2009b). Simplified Facial Animation Control Utilizing Novel Input Devices: A Comparative Study. In *International Conference on Intelligent User Interfaces (IUI '09)*, pages 197–206.
- Bee, N., Franke, S., and André, E. (2009c). Relations between facial display, eye gaze and head tilt: Dominance perception variations of virtual agents. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–7. IEEE.
- Bee, N., Pollock, C., André, E., and Walker, M. (2010a). Bossy or Wimpy: Expressing Social Dominance by Combining Gaze and Linguistic Behaviors. In Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., and Safonova, A., editors, *Intelligent Virtual Agents*, volume 6356 of *Lecture Notes in Computer Science*, pages 265–271, Berlin, Heidelberg. Springer Berlin / Heidelberg.
- Bee, N., Prendinger, H., André, E., and Ishizuka, M. (2006a). Automatic preference detection by analyzing the gaze 'cascade effect'. In *Proceedings 2nd COGAIN Annual Conference on Communication by Gaze Interaction*, pages 63–66.

- Bee, N., Prendinger, H., Nakasone, A., André, E., and Ishizuka, M. (2006b). AutoSelect: What You Want Is What You Get: Real-Time Processing of Visual Attention and Affect. In *Perception and Interactive Technologies (PIT)*, pages 40–52.
- Bee, N., Wagner, J., André, E., Vogt, T., Charles, F., Pizzi, D., and Cavazza, M. (2010b). Discovering eye gaze behavior during human-agent conversation in an interactive storytelling application. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction, ICMI-MLMI '10*, New York, NY, USA. ACM.
- Bente, G., Eschenburg, F., and Aelker, L. (2007). Effects of simulated gaze on social presence, person perception and personality attribution in avatar-mediated communication. In *PRESENCE 2007*.
- Bossi, J. (1995). *Augen-Blicke. Zur Psychologie des Flirts*. Huber, Bern.
- Cassell, J. and Bickmore, T. W. (2003). Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Model. User-Adapt. Interact.*, 13(1-2):89–132.
- Cavazza, M., Lugrin, J. L., Pizzi, D., and Charles, F. (2007). Madame bovary on the holodeck: immersive interactive storytelling. In *Proceedings of the 15th international conference on Multimedia, MULTIMEDIA '07*, pages 651–660, New York, NY, USA. ACM.
- Cavazza, M., Pizzi, D., Charles, F., Vogt, T., and André, E. (2009). Emotional input for character-based interactive storytelling. In *AAMAS '09: Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems*, pages 313–320, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.
- Chang, E. and Jenkins, O. C. (2006). Sketching articulation and pose for facial animation. In *Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation, SCA '06*, pages 271–280, Aire-la-Ville, Switzerland, Switzerland. Eurographics Association.
- Cohen, D. (1992). *Body Language in Relationships*. Sheldon Press.
- Colburn, A., Cohen, M., and Drucker, S. (2000). The Role of Eye Gaze in Avatar Mediated Conversational Interfaces. Technical report, Microsoft Research.
- Courgeon, M., Martin, J. C., and Jacquemin, C. (2008). User's gestural exploration of different virtual agents' expressive profiles. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems - Volume 3*,

- AAMAS '08, pages 1237–1240, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.
- Cowie, R., Cowie, E. D., Savvidou, S., McMahon, E., Sawey, M., and Schröder, M. (2000). FEELTRACE: an instrument for recording perceived emotion in real time.
- Damian, I. (2011). Customizing agent interactions. Master's thesis, Augsburg University, Institute for Computer Science.
- Dow, S., Mehta, M., Harmon, E., MacIntyre, B., and Mateas, M. (2007). Presence and engagement in an interactive drama. In *Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '07*, pages 1475–1484, New York, NY, USA. ACM.
- Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34(4):455–470.
- Duchowski, A. T. (2007). *Eye Tracking Methodology: Theory and Practice*. Springer, 2nd edition.
- Eichner, T., Prendinger, H., André, E., and Ishizuka, M. (2007). Attentive Presentation Agents. In *Intelligent Virtual Agents (IVA 2007)*, pages 283–295.
- Ekman, P. and Friesen, W. (1975). *Unmasking the Face*. Prentice Hall.
- Ekman, P. and Friesen, W. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto.
- Ellsworth, P. C. and Ludwig, L. M. (1972). Visual behavior in social interaction. *Journal of Communication*, 22:375–403.
- Eysenck, H. J. and Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. London: Hodder and Stoughton.
- Facial Expression Repertoire (2008). Filmakademie Baden-Württemberg. <http://research.animationsinstitut.de/>.
- Flaubert, G. (1856). *La revue de Paris*. France.
- Frydrych, M., Dobsik, M., Kätsyri, J., and Sams, M. (2003). Toolkit for animation of finnish talking head. In *ISCA Tutorial and Research Workshop on Audio Visual Speech Processing (AVSP'03)*, pages 199–204. International Speech Communication Association.

- Fukayama, A., Ohno, T., Mukawa, N., Sawaki, M., and Hagita, N. (2002). Messages embedded in gaze of interface agents — impression management with agent's gaze. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 41–48, New York, NY, USA. ACM Press.
- Garau, M., Slater, M., Bee, S., and Sasse, M. A. (2001). The impact of eye gaze on communication using humanoid avatars. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 309–316, New York, NY, USA. ACM Press.
- Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., and Sasse, A. M. (2003). The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 529–536, New York, NY, USA. ACM Press.
- Gillies, M. F. P. and Dodgson, N. A. (2002). Eye movements and attention for behavioural animation. *The Journal of Visualization and Computer Animation*, 13(5):287–300.
- Givens, D. B. (1978). The nonverbal basis of attraction: flirtation, courtship, and seduction. *Psychiatry*, 41(4):346–359.
- Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Journal of Personality and Social Psychology*, 59(6):1216–1229.
- Gosling, S. D., Rentfrow, P. J., and Jr (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6):504–528.
- Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., van der Werf, R., and Morency, L. P. (2006). Virtual Rapport. In Gratch, J., Young, M., Aylett, R., Ballin, D., and Olivier, P., editors, *Intelligent Virtual Agents (IVA 2006)*, volume 4133, pages 14–27, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Hall, E. T. (1963). A system for notation of proxemic behavior. *American Anthropologist*, 65:1003–1026.
- Hansen, J. P., Hansen, D. W., and Johansen, A. S. (2001). Bringing Gaze-based Interaction Back to Basics. *Systems, Social and Internationalization Design Aspects of Human-computer Interaction*.
- Häring, M., Bee, N., and André, E. (2011). Creation and evaluation of emotion expression with body movement, sound and eye color for humanoid robots. In *IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man)*.

- Isokoski, P. (2000). Text input methods for eye trackers using off-screen targets. In *ETRA '00: Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 15–21, New York, NY, USA. ACM.
- Jackson, C. J. (2001). Comparison between eysenck's and gray's models of personality in the prediction of motivational work criteria. *Personality and Individual Differences*, 31(2):129–144.
- Jacob, R. J. K. (1991). The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems*, 9(2):152–169.
- Jacquemin, C. (2007). Pogany: A tangible cephalomorphic interface for expressive facial animation. In Paiva, A., Prada, R., and Picard, R., editors, *Affective Computing and Intelligent Interaction*, volume 4738 of *Lecture Notes in Computer Science*, pages 558–569, Berlin, Heidelberg. Springer Berlin / Heidelberg.
- Kätsyri, J. (2006). *Human Recognition of Basic Emotions from Posed and Animated Dynamic Facial Expressions*. PhD thesis, Helsinki University of Technology.
- Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologica*, 26:22–63.
- Khullar, S. C. and Badler, N. I. (2001). Where to Look? Automating Attending Behaviors of Virtual Human Characters. *Autonomous Agents and Multi-Agent Systems*, 4(1-2):9–23.
- Kipp, M. and Gebhard, P. (2008). IGaze: Studying Reactive Gaze Behavior in Semi-immersive Human-Avatar Interactions. In *Intelligent Virtual Agents (IVA '08)*, pages 191–199.
- Klinke, C. L. (1986). Gaze and Eye Contact: A Research Review. *Psychological Bulletin*, 100(1):78–100.
- Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, 20(3):165–182.
- Költringer, T., Van, M. N., and Grechenig, T. (2007). Game controller text entry with alphabetic and multi-tap selection keyboards. In *CHI '07: CHI '07 extended abstracts on Human factors in computing systems*, pages 2513–2518, New York, NY, USA. ACM.
- Lance, B. and Marsella, S. (2007). Emotionally Expressive Head and Body Movement During Gaze Shifts. In Pelachaud, C., Martin, J.-C., André, E., Chollet,

- G., Karpouzis, K., and Pelé, D., editors, *Intelligent Virtual Agents*, volume 4722 of *Lecture Notes in Computer Science*, chapter 8, pages 72–85. Springer Berlin / Heidelberg, Berlin, Heidelberg.
- Lance, B. and Marsella, S. (2008). The Relation between Gaze Behavior and the Attribution of Emotion: An Empirical Study. In Prendinger, H., Lester, J., and Ishizuka, M., editors, *Intelligent Virtual Agents (IVA '08)*, volume 5208 of *Lecture Notes in Computer Science*, pages 1–14, Berlin, Heidelberg. Springer Berlin / Heidelberg.
- Leathers, D. G. (1991). *Successful Nonverbal Communication: Principles and Applications*. Macmillan Pub Co.
- Lee, S. P., Badler, J. B., and Badler, N. I. (2002). Eyes alive. *ACM Transactions on Graphics*, 21(3):637–644.
- Lewis, J. P., Mooser, J., Deng, Z., and Neumann, U. (2005). Reducing blendshape interference by selected motion attenuation. In *Proceedings of the 2005 symposium on Interactive 3D graphics and games, I3D '05*, pages 25–29, New York, NY, USA. ACM.
- MacKenzie, S. I. (2003). Motor Behaviour Models for Human-Computer Interaction. In Carroll, J. M., editor, *HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science*.
- Majoranta, P., Aula, A., and Räihä, K. J. (2004). Effects of feedback on eye typing with a short dwell time. In *ETRA '04: Proceedings of the 2004 symposium on Eye tracking research & applications*, pages 139–146, New York, NY, USA. ACM.
- Majoranta, P., MacKenzie, I., Aula, A., and Räihä, K. J. (2006). Effects of feedback and dwell time on eye typing speed and accuracy. *Universal Access in the Information Society*, 5(2):199–208.
- McCloud, S. (2006). *Making Comics: Storytelling Secrets of Comics, Manga and Graphic Novels*. Harper Paperbacks.
- McCrae, R. R. and Costa, P. T. (1987). Validation of a five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52:81–90.
- McCrae, R. R. and John, O. P. (1992). An introduction to the Five-Factor model and its applications. *Journal of Personality*, 60(2):175–215.

- Mehrabian, A. (1995). Framework for a comprehensive description and measurement of emotional states. *Genetic, social, and general psychology monographs*, 121(3):339–361.
- Mehrabian, A. and Russell, J. A. (1974). *An Approach to Environmental Psychology*. The MIT Press.
- Morency, L. P., Christoudias, M. C., and Darrell, T. (2006). Recognizing gaze aversion gestures in embodied conversational discourse. In *ICMI '06: Proceedings of the 8th international conference on Multimodal interfaces*, pages 287–294, New York, NY, USA. ACM Press.
- Nakano, Y. I., Reinstein, G., Stocky, T., and Cassell, J. (2003). Towards a model of face-to-face grounding. In *ACL '03: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pages 553–561, Morristown, NJ, USA. Association for Computational Linguistics.
- Nakano, Y. I. and Yamaoka, Y. (2009). Information state based multimodal dialogue management: Estimating conversational engagement from gaze information. In Ruttkay, Z., Kipp, M., Nijholt, A., and Vilhjálmsón, H. H., editors, *Intelligent Virtual Agents, 9th International Conference, IVA 2009, Amsterdam, The Netherlands, September 14-16, 2009, Proceedings*, volume 5773 of *Lecture Notes in Computer Science*, pages 531–532. Springer.
- Nataneli, G. and Faloutsos, P. (2007). Sketching facial expressions. In *ACM SIGGRAPH 2007 sketches*, SIGGRAPH '07, New York, NY, USA. ACM.
- Oat, C. (2007). Animated wrinkle maps. In *ACM SIGGRAPH 2007 courses*, SIGGRAPH '07, pages 33–37, New York, NY, USA. ACM.
- Ochs, M., Niewiadomski, R., Pelachaud, C., and Sadek, D. (2005). Intelligent expressions of emotions. In *Affective Computing and Intelligent Interaction (ACII)*.
- Pan, X. and Slater, M. (2007). A Preliminary Study of Shy Males Interacting with a Virtual Female. In *PRESENCE 2007: The 10th Annual International Workshop on Presence*.
- Pandzic, I. S. and Forchheimer, R., editors (2003). *MPEG-4 Facial Animation: The Standard, Implementation and Applications*. John Wiley & Sons, Inc., New York, NY, USA.
- Pasquariello, S. and Pelachaud, C. (2001). Greta: A simple facial animation engine. In *6th Online World Conference on Soft Computing in Industrial Applications*.

- Pelachaud, C. and Bilvi, M. (2003). Modelling gaze behavior for conversational agents. In *Intelligent Virtual Agents (IVA)*.
- Perlin, K. (1998). Quikwriting: continuous stylus-based text entry. In *UIST '98: Proceedings of the 11th annual ACM symposium on User interface software and technology*, pages 215–216, New York, NY, USA. ACM Press.
- Peters, C., Pelachaud, C., Bevacqua, E., Mancini, M., and Poggi, I. (2005). A model of attention and interest using Gaze behavior. In *Intelligent Virtual Agents (IVA'05)*, pages 229–240, London, UK. Springer-Verlag.
- Pfeiffer, T. and Latoschik, M. E. (2004). Resolving object references in multimodal dialogues for immersive virtual environments. In *Virtual Reality, 2004. IEEE*.
- Pfeiffer, T., Latoschik, M. E., and Wachsmuth, I. (2009). Evaluation of binocular eye trackers and algorithms for 3d gaze interaction in virtual reality environments. *Journal of Virtual Reality and Broadcasting*, 16(5).
- Prasov, Z. and Chai, J. Y. (2008). What's in a Gaze? The Role of Eye-Gaze in Reference Resolution in Multimodal Conversational Interfaces. In *ACM 12th International Conference on Intelligent User Interfaces (IUI)*.
- Rehm, M. and André, E. (2005). From chatterbots to natural interaction - Face to face communication with Embodied Conversational Agents. *IEICE Transactions on Information and Systems, Special Issue on Life-Like Agents and Communication*, 88-D(11):2445–2452.
- Russell, J. A. and Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3):273–294.
- Ruttkay, Z., Noot, H., and Hagen, P. (2003). Emotion disc and emotion squares: Tools to explore the facial expression space. *Computer Graphics Forum*, 22.
- Sagar, M. (2006). Facial performance capture and expressive translation for King Kong. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Sketches*, New York, NY, USA. ACM.
- Salvucci, D. D. and Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *ETRA '00: Proceedings of the symposium on Eye tracking research & applications*, pages 71–78, New York, NY, USA. ACM Press.
- Sander, D., Grandjean, D., Kaiser, S., Wehrle, T., and Scherer, K. R. (2007). Interaction effects of perceived gaze direction and dynamic facial expression: Evidence for appraisal theories of emotion. *European Journal of Cognitive Psychology*, 19(3):470–480.

- Schmidt, A. (2000). Implicit human computer interaction through context. *Personal and Ubiquitous Computing*, 4(2):191–199.
- Schulz, N. (2008). Rendering expressive faces with Horde3D. Master's thesis, Augsburg University, Institute for Computer Science.
- Shimojo, S., Simion, C., Shimojo, E., and Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nat Neurosci*, 6(12):1317–1322.
- Sidner, C. L., Kidd, C. D., Lee, C., and Lesh, N. (2004). Where to look: a study of human-robot engagement. In *IUI '04: Proceedings of the 9th international conference on Intelligent user interfaces*, pages 78–84, New York, NY, USA. ACM Press.
- Spakov, O. and Miniotos, D. (2004). On-line adjustment of dwell time for target selection by gaze. In *NordiCHI '04: Proceedings of the third Nordic conference on Human-computer interaction*, pages 203–206, New York, NY, USA. ACM.
- Spencer-Smith, J., Wild, H., Innes-Ker, A. H., Townsend, J., Duffy, C., Edwards, C., Ervin, K., Merritt, N., and Paik, J. W. (2001). Making faces: creating three-dimensional parameterized models of facial expression. *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc*, 33(2):115–123.
- Starker, I. and Bolt, R. A. (1990). A gaze-responsive self-disclosing display. In *CHI '90: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 3–10, New York, NY, USA. ACM.
- Steptoe, W., Wolff, R., Murgia, A., Guimaraes, E., Rae, J., Sharkey, P., Roberts, D., and Steed, A. (2008). Eye-tracking for avatar eye-gaze and interactional analysis in immersive collaborative virtual environments. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work, CSCW '08*, pages 197–200, New York, NY, USA. ACM.
- Suontphunt, T., Mo, Z., Neumann, U., and Deng, Z. (2008). Interactive 3D facial expression posing through 2D portrait manipulation. In *Proceedings of graphics interface 2008, GI '08*, pages 177–184, Toronto, Ont., Canada, Canada. Canadian Information Processing Society.
- Thalmann, D. (1993). Using Virtual Reality Techniques in the Animation Process. *Virtual Reality Systems*, pages 143–159.
- Tramitz, C. (1992). *Auf den ersten Blick*. ADMOS Media GmbH.

- Urbina, M. H. and Huckauf, A. (2007). Dwell time free eye typing approaches. In *COGAIN 2007: Gaze-based Creativity and Interacting with Games and On-line Communities*.
- Vertegaal, R., Slagter, R., van der Veer, G., and Nijholt, A. (2001). Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In *Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '01*, pages 301–308, New York, NY, USA. ACM.
- Vinayagamoorthy, V., Garau, M., Steed, A., and Slater, M. (2004). An Eye Gaze Model for Dyadic Interaction in an Immersive Virtual Environment: Practice and Experience. *Computer Graphics Forum*, 23(1):1–11.
- Vogt, T., André, E., and Bee, N. (2008). EmoVoice - A framework for online recognition of emotions from voice. In *Proceedings of Workshop on Perception and Interactive Technologies*, pages 188–199. Springer.
- Wagner, J., André, E., and Jung, F. (2009). Smart Sensor Integration: A Framework for Multimodal Emotion Recognition in Real-Time. In *International Conference on Affective Computing & Intelligent Interaction (ACII)*, pages 209–216. IEEE.
- Ward, D. J. and MacKay, D. J. C. (2002). Fast Hands-free Writing by Gaze Direction.
- Watson, D. and Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin*, 96(3):465–490.
- Wobbrock, J. O., Rubinstein, J., Sawyer, M. W., and Duchowski, A. T. (2008). Longitudinal Evaluation of Discrete Consecutive Gaze Gestures for Text Entry. In *ETRA '08: Proceedings of the 2006 symposium on Eye tracking research & applications*.

CURRICULUM VITAE

NIKOLAUS BEE

<http://hcm-lab.de/nikolaus-bee.html>

WORK EXPERIENCE

- since 11/2011 **Researcher**, BMW Research and Technology, Munich, Germany
- 11/2006-10/2011 **Research Assistant**, Lab for *Human-Centered Multimedia*, Augsburg University, Germany
- 10/2003-04/2005 **Student Researcher**, Lab for *Human-Centered Multimedia*, Augsburg University, Germany
Working for *HUMAINE* (Human-Machine Interaction Network on Emotion), an EU-funded Network of Excellence.
- 08/2003-10/2003 **Internship**, T-Systems GEI GmbH, Ulm, Germany
Topic: Analysis and Design of a Component-Based Remote Maintenance System
- 01/2001-07/2003 **Student Researcher**, Chair of *Software Engineering*, Augsburg University, Germany

STAY ABROAD

- 11/2005-08/2006 **National Institute of Informatics (NII)**, Tokyo, Japan
Research Topic: Real-Time System for Multimodal Fusion of User Behavior for Emotion and Attention Recognition

EDUCATION

- 11/2006-01/2013 **Augsburg University**: PhD student in Computer Science,
Topic: Affective and Attentive Interaction with Virtual Humans in Gaze-based Settings
- 10/2000-03/2006 **Augsburg University**: Diploma student in Computer Science
(minors: Human-Machine Interaction, Software Engineering, Economics), Grade: 1.9
Diploma thesis: *Akquisition von sprachlichen und physiologischen Daten zur mehrkanaligen Emotionsanalyse* [Acquisition of Speech and Physiological Data for Multimodal Emotional Analysis]
Supervisor: Prof. Dr. Elisabeth André, Grade: 1.0

RESEARCH TOPICS

Human-Machine Interaction, Real-Time Systems, Pattern Recognition, Machine Learning

Measuring and Recognition of User Behavior through Gaze, Speech, Physiology and Facial Expressions

Interacting with Virtual Worlds with Different Interaction Devices (Wiimote, Smartphones, iPad, Eye Tracker, Face Tracker, Speech, Data Glove, Physiological Sensors, Kinect)

RESEARCH PROJECTS

2008-2011	DynaLearn (EU-funded STREP)
2008-2011	IRIS: Integrating Research in Interactive Storytelling (EU-funded Network of Excellence)
2006-2008	CUBE-G: CUlture-adaptive BEhavior Generation for interactions with embodied conversational agents (DFG-funded)
2006-2010	CALLAS: Conveying Affectiveness in Leading-edge Living Adaptive Systems (EU-funded IP)
2004-2007	HUMAINE: Human-Machine Interaction Network on Emotion (EU-funded Network of Excellence)

AWARDS

GALA (Gathering of Animated Lifelike Agents) Award in conjunction with the International Conference on Intelligent Virtual Agents (IVA) 2006 (Public and Jury Award) and supervising of winners in 2008 (Jury Award) and 2009 (Public Award)

<http://hmi.ewi.utwente.nl/gala/>

"Best Paper" nomination at the International Conference on Intelligent Virtual Agents (IVA) in 2009

PROFESSIONAL SERVICE

Program Committees:

- International Joint Conference on Artificial Intelligence (IJ-CAI 2011)
- Workshop on Eye Gaze in Intelligent Human Machine Interaction in conjunction with the international conference on Intelligent User Interfaces (2010, 2011)

Conference Reviews:

- Conference on Human Factors in Computing Systems (CHI 2011)
- Conference on Human-Computer Interaction (INTERACT 2011)
- International Joint Conference on Artificial Intelligence (IJ-CAI 2011)
- Workshop on Eye Gaze in Intelligent Human Machine Interaction (2010, 2011)
- Conference on Designing Interactive Systems (DIS 2010)
- Conference on Intelligent User Interfaces (IUI 2007, 2010)
- Conference on Computer Animation and Social Agents (CASA 2010)
- Annual Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB 2008)
- International Symposium in Robot and Human Interactive Communication (RO-MAN 2007)

Journal Reviews:

- ACM Transactions on Interactive Intelligent Systems
- Journal of Autonomous Agents and Multi-Agent Systems

TEACHING

Practical Courses:

Emotion Recognition (2003), Pattern Recognition (2004), Interaction with Virtual Worlds (2007), Interaction with Virtual Agents (2008), Game Programming (2009 and 2010), Multimodal User Interfaces (2010)

Lectures:

Introduction to Game Programming (2008, 2009, 2010)

Seminars:

Attentive Agents and Technologies (2007), Interactive Multimedia Systems (2008), Game Programming (2009)

Theses:

Supervision of 14 Bachelor (BSc) theses and 3 Master (MSc) theses

ADDITIONAL**Languages:**

German (native), English (fluent), French (basic), Japanese (basic)

Memberships:

Gesellschaft für Informatik (GI), Association for Computing Machinery (ACM)

Public Relations:

CineUni (2011), 40th anniversary of the Augsburg University (2010), DW-TV (2010), Open Lab Day (2004, 2007, 2008, 2009), Information day for students (2009, 2010), CeBIT (2007), NII Open House (2006)

Programming Skills:

C/C++, Java, C#, SPSS, Matlab, Python, Perl

PUBLICATIONS

- 2012** J. Wagner, F. Lingenfelser, N. Bee, E. André. Social Signal Interpretation Real-time Sensing of Affective and Social Signals, *Künstliche Intelligenz: Special Issue on Emotional Computing*, 2012.
- 2011** M. Häring, N. Bee and E. André. Creation and Evaluation of Emotion Expression with Body Movement, Sound and Eye Color for Humanoid Robots. *IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man)*, 2011.
- I. Damian, B. Endrass, N. Bee, E. André. Individualized Agent Interactions. *International Conference on Intelligent Virtual Agents (IVA)*, 2011.
- K. Leichtenstern, N. Bee, E. André, U. Berkmüller and J. Wagner. Physiological Measurement of Trust-Related Behavior in Trust-Neutral and Trust-Critical Situations. *IFIP International Conference on Trust Management (IFIPTM)*, 2011.
- D. Arellano, N. Bee, K. Janowski, E. André, J. Varona, F. J. Perales. Influence of Head Orientation in Perception of Personality

Traits in Virtual Agents. *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2011.

2010

N. Bee, J. Wagner, E. André, T. Vogt, F. Charles, D. Pizzi and M. Cavazza. Discovering Eye Gaze Behavior during Human-Agent Conversation in an Interactive Storytelling Application. *International Conference on Multimodal Interfaces and 7th Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI)*, 2010.

N. Bee, C. Pollock, E. André and M. Walker. Bossy or Wimpy: Expressing Social Dominance by Combining Gaze and Linguistic Behaviors. *International Conference on Intelligent Virtual Agents (IVA)*, 2010.

N. Bee, J. Wagner, E. André, F. Charles, D. Pizzi and M. Cavazza. Multimodal Interaction with a Virtual Character in Interactive Storytelling. *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2010.

N. Bee, J. Wagner, E. André, T. Vogt, F. Charles, D. Pizzi and M. Cavazza. Gaze Behavior during Interaction with a Virtual Character in Interactive Storytelling. *AAMAS 2010 Workshop on Interacting with ECAs as Virtual Characters*, 2010.

N. Bee, E. André, T. Vogt and P. Gebhard. *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues*, Y. Wilks, Eds., John Benjamins, 2010, ch. The use of affective and attentive cues in an empathic computer-based companion, pp. 131-142.

N. Bee, J. Wagner, E. André, F. Charles, D. Pizzi and M. Cavazza. Interacting with a Gaze-Aware Virtual Character. *International Workshop on Eye Gaze in Intelligent Human Machine Interaction (IUI)*, 2010.

2009

N. Bee, E. André and S. Tober. Breaking the Ice in Human-Agent Communication: Eye-Gaze Based Initiation of Contact with an Embodied Conversational Agent. *International Conference on Intelligent Virtual Agents (IVA)*, 2009.

N. Bee, B. Falk and E. André. Simplified Facial Animation Control Utilizing Novel Input Devices: A Comparative Study. *International Conference on Intelligent User Interfaces (IUI)*, 2009, pp. 197-206.

N. Bee, S. Franke and E. André. Relations between Facial Display, Eye Gaze and Head Tilt: Dominance Perception Variations of Virtual Agents. *Affective Computing and Intelligent Interaction (ACII)*, 2009.

B. Endrass, M. Boegler, N. Bee and E. André. What Would You Do in their Shoes? Experiencing Different Perspectives in an In-

teractive Drama for Multiple Users. *International Conference on Interactive Digital Storytelling (ICIDS)*, 2009, pp. 258-268.

M. Wissner, N. Bee, J. Kienberger and E. André. To See and to Be Seen in the Virtual Beer Garden – A Gaze Behavior System for Intelligent Virtual Agents in a 3D Environment. *Advances in Artificial Intelligence, Annual Conference on AI (KI)*, 2009, pp. 500-507.

M. Rehm, Y. Nakano, E. André, T. Nishida, N. Bee, B. Endrass, M. Wissner, A. A. Lipi and H. Huang. From observation to simulation: generating culture-specific behavior for interactive systems. *AI & Society*, vol. 24, no. 3, pp. 267-280, 2009.

M. Rehm, E. André, N. Bee, B. Endrass, M. Wissner, Y. Nakano, A. A. Lipi, T. Nishida and H. Huang. *Multimodal Corpora: From Models of Natural Interaction to Systems and Applications*, M. Kipp, J. Martin, P. Paggio and D. Heylen, Eds., Springer, 2009, ch. Creating Standardized Video Recordings of Multimodal Interactions Across Cultures, pp. 138-159.

N. Bee, E. André, T. Vogt and P. Gebhard. First ideas on the use of affective cues in an empathic computer-based companion. *AAMAS '09 Workshop on Empathic Agents*, 2009.

H. Prendinger, A. Hyrskykari, M. Nakayama, H. Istance, N. Bee and Y. Takahasi. Attentive interfaces for users with disabilities: eye gaze for intention and uncertainty estimation. *Journal of Universal Access in the Information Society*, March 2009.

2008

N. Bee and E. André. Writing with Your Eye: A Dwell Time Free Writing System Adapted to the Nature of Human Eye Gaze, *Workshop on Perception and Interactive Technologies for Speech-Based Systems*, 2008.

N. Bee and E. André. Cultural gaze behavior to improve the appearance of virtual agents. *IUI Workshop on Enculturating Interfaces (ECI)*, 2008.

M. Rehm, N. Bee and E. André. Wave Like an Egyptian: Accelerometer Based Gesture Recognition for Culture Specific Interactions. *HCI 2008 Culture, Creativity, Interaction*, 2008.

M. Rehm, T. Vogt, M. Wissner and N. Bee. Dancing the Night Away – Controlling a Virtual Karaoke Dancer by Multimodal Expressive Cues. *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2008.

T. Vogt, E. André and N. Bee. EmoVoice - A framework for online recognition of emotions from voice. *Workshop on Perception and Interactive Technologies for Speech-Based Systems*, 2008.

- 2007**
- F. Charles, S. Lemercier, T. Vogt, N. Bee, M. Mancini, J. Urbain, M. Price, E. André, C. Pelachaud and M. Cavazza. Affective Interactive Narrative in the CALLAS Project. *International Conference on Virtual Storytelling*, 2007.
- A. Hoekstra, H. Prendinger, N. Bee, D. Heylen, M. Ishizuka. Highly realistic 3D presentation agents with visual attention capability. *International Symposium on Smart Graphics (SG-07)*, 73-84, 2007.
- M. Rehm, N. Bee, B. Endrass, M. Wissner and E. André. Too close for comfort? Adapting to the user's cultural background. *International Workshop on Human-Centered Multimedia*, 2007.
- M. Rehm, E. André, N. Bee, B. Endrass, M. Wissner, Y. Nakano, T. Nishida and H. Huang. The CUBE-G approach – Coaching culture-specific nonverbal behavior by virtual agents. *Conference of the International Simulation and Gaming Association (ISAGA)*, 2007.
- M. Rehm, N. Bee and B. Endrass. Increasing Cultural Awareness by Games with Embodied Conversational Agents. *Learning with Games*, 2007.
- A. Hoekstra, H. Prendinger, N. Bee, D. Heylen, M. Ishizuka. Presentation Agents That Adapt to Users' Visual Interest and Follow Their Preferences. *International Conference on Computer Vision Systems (ICVS 2007)*, 2007.
- 2006**
- H. Prendinger, A. Hoekstra, N. Bee, M. Nischt and M. Ishizuka. Visual interest contingent presentation agents. *Joint Agent Workshop & Symposium (JAWS 2006)*, 2006.
- N. Bee, H. Prendinger, E. André, M. Ishizuka. Automatic preference detection by analyzing the gaze 'cascade effect'. *COGAIN Annual Conference on Communication by Gaze Interaction*, 63-66, 2006.
- N. Bee, H. Prendinger, A. Nakasone, E. André and M. Ishizuka. AutoSelect: What You Want Is What You Get, Real-Time Processing of Visual Attention and Affect. *Perception and Interactive Technologies (PIT)*, 40-52, 2006.
- 2004**
- J. Kim, N. Bee, J. Wagner, E. André. Emote to Win: Affective Interactions with a Computer Game Agent. *GI-Edition - Lecture Notes in Informatics (LNI)*, 159-164, 2004.