

Believable Virtual Characters in Human-Computer Dialogs

Yvonne Jung¹, Arjan Kuijper^{1,2}, Dieter Fellner^{1,2}, Michael Kipp³, Jan Miksatko³, Jonathan Gratch⁴, and Daniel Thalmann⁵

¹Technische Universität Darmstadt / Fraunhofer IGD, Germany

²Institut für Computer Graphik & Wissensvisualisierung, TU Graz, Austria

³Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Germany

⁴University of Southern California, CA, USA

⁵Institute for Media Innovation, Nanyang Technological University, Singapore

Abstract

For many application areas, where a task is most naturally represented by talking or where standard input devices are difficult to use or not available at all, virtual characters can be well suited as an intuitive man-machine-interface due to their inherent ability to simulate verbal as well as nonverbal communicative behavior. This type of interface is made possible with the help of multimodal dialog systems, which extend common speech dialog systems with additional modalities just like in human-human interaction. Multimodal dialog systems consist at least of an auditive and graphical component, and communication is based on speech and nonverbal communication alike. However, employing virtual characters as personal and believable dialog partners in multimodal dialogs entails several challenges, because this requires not only a reliable and consistent motion and dialog behavior but also regarding nonverbal communication and affective components. Besides modeling the “mind” and creating intelligent communication behavior on the encoding side, which is an active field of research in artificial intelligence, the visual representation of a character including its perceivable behavior, from a decoding perspective, such as facial expressions and gestures, belongs to the domain of computer graphics and likewise implicates many open issues concerning natural communication. Therefore, in this report we give a comprehensive overview how to go from communication models to actual animation and rendering.

Categories and Subject Descriptors (according to ACM CCS): H.5.1 [Information Interfaces and Presentation (e.g., HCI)]: Multimedia Information Systems—Artificial, augmented, and virtual realities; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation

1. Introduction

During the past few years there has been an increasing interest in virtual characters [GVT08], not only in Virtual Reality (VR), computer games or online communities such as Second Life, but also for dialog-based systems like tutoring systems or edu- and infotainment applications. This is directly associated with the major challenges of Human-Computer-Interface technologies in general [PI04] and immersive Augmented and Virtual Reality concepts in particular, as they are both aimed at developing intuitive interfaces instead of the standard GUI interaction style (i.e. WIMP), which basically has not changed for more than three decades. However, since computing power becomes more and more ubiquitous, it is

inevitable to extend these traditional interaction methods. In this regard, virtual characters are well suited as an intuitive interface by simulating verbal as well as nonverbal communicative behavior (as shown in Figure 1), for ensuring intuitive interactions even for inexperienced users and beyond standard settings.

Possible fields of application embrace situations, where a task is most naturally represented by talking, like in interpersonal skills education (cf. e.g. [JRS*07]), or where typical input devices like mouse and keyboard are difficult to use or not available at all, like in immersive VR settings or in Augmented Reality (AR) supported on-site manuals for maintenance scenarios [BS06]. Other examples are assistance sys-



Figure 1: Example of nonverbal communication: on the left a screenshot of a TV talk show is shown and on the right a comic-style virtual character that is gesticulating similarly.

tems like interactive manuals or virtual tour guides, where the virtual human explains cultural heritage sites or the usage of a new device. Somewhat unsuitable areas for conversational interfaces are for instance typical office applications like wordprocessing and applications that are mainly driven by direct manipulation tasks such as painting or moving around photos or maps on a multi-touch table. However, such gestural input is just another mode of communication and thereby part of a multimodal interface system, which therefore needs to account for both, multimodal input recognition and output generation.

This report focuses on the visualization component of multimodal dialog systems (MDS, where we use this term as defined by Wahlster [Wah06]), which in this context is also known as the ‘surface realizer’ [LPR07, HK09] and mainly deals with the graphical realization of the embodied agent including its nonverbal output. In principle, this defines the requirements for such a component but also the scope of this report. Since in this context the virtual humans mostly talk and gesticulate, we have a reduction of animation complexity, since locomotion and other intricate movements as well as path planning aspects are of minor importance. Furthermore, in dialog systems only one or a few characters are used. Therefore, e.g. crowd rendering [TM07] and the like does not necessarily need to be considered.

Though there already exist many systems to simulate virtual characters, they are mostly focused on certain subdomains, designed as standalone application using proprietary formats and in-house libraries, or they do not address the demands of interactive, dynamic environments as particularly given in Mixed Reality (MR) environments [Tüm07]. Likewise, in [VGS*06] the authors concluded that “*this is a diverse area of research and this diversity of research is itself a challenge to researchers in this field: each character system has been designed to investigate a particular aspect of non-verbal communication. It is not yet clear if all of this research can be integrated into a single platform [...]*.” However, with recent developments in character animation and emerging 3D standards like Collada [AB06] and X3D [Web08] on the one hand, as well as component-based and service-oriented system architectures and unified

interface languages like BML, as in SAIBA [VCC*07], on the other, this goal now comes into reach. One challenge thus is finding generic frameworks for interactive agents (cp. Greta [NBMP09, Pel05] or SAIBA), including appropriate high-level interfaces and standardizable control languages like FML and BML, for specifying, coordinating, and synchronizing the communicative behavior of virtual humans, discussed in sections 2 and 3.

The prospects of advanced real-time rendering techniques due to the rapid GPU development cycles still are mostly ignored in embodied agents research. For example in the SmartKom architecture [HR06], the whole presentation component makes at most ten percent of the complete system design, where the modality-specific output is handled by the character animation component that realizes the perceivable behavior of the embodied conversational agent (ECA). In addition, the agent needs to be tightly integrated into larger applications to allow for interactions between the virtual character, the user, and the 3D world. Also, from a decoding perspective, the environment can be used for communication too. Therefore, character-external factors like lighting and camera control are surveyed in sections 6 and 7.

Embedding the character is also necessary to avoid missing contextual relevance or interactions with the ECA that appear artificial [SOM*10]. It is thus necessary to embed all techniques into a complete system – not only to simplify the integration of virtual characters into whole 3D applications, but also to ease the interaction between real and virtual humans. This implies having building blocks for gestures, speech, and emotions, as well as adequate layers of abstraction that enable system internal and external use through a unified interface [GRA*02, HK09]. The presentation component of an MDS hence needs to be able to integrate relevant functionalities (which will be explained within the course of this report) and provide them in a manageable manner [JB08, KHGS10].

To also account for availability, efficiency, and sustainability, we further discuss possibilities to integrate those techniques into suitable and established standards. One such example is the open ISO standard X3D, which is currently the only standardized 3D deployment format. Besides this, related subquestions such as camera and animation control are considered too, since a really interactive agent requires a high degree of control over its body, face, voice, and its physiological or externally observable processes in general, whereas the system must be able to clarify or emphasize any of these with suitable camera work [CO09].

Another focus lies on the graphical representation of virtual characters with particular emphasis on the dynamic aspects of rendering. Therefore, relevant building blocks for rendering and animation are discussed, which not only provide flexible control concerning gestures, facial expression, and speech, but also consider resultant dependencies that need to be simulated during runtime, such as long hair blow-

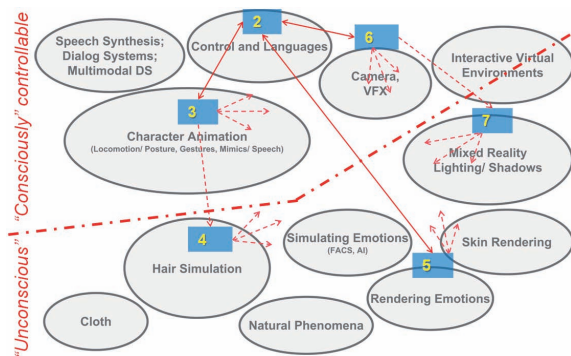


Figure 2: Coverage of fields of research by interconnected components (shown in blue, the numbers refer to the sections) with respect to related work (shown as bubbles).

ing in the wind or tears dripping down. This also includes psycho-physiological processes like blushing and crying. Real humans cannot control them deliberately, but for characters they need to be consistent and synchronous with their motor response. Though these effects until recently were mostly ignored in research [JK06], they are essential for the correct perception of strong emotions in the context of non-verbal communication.

Therefore, models that can be used for the automatic synthesis of realistic communicative behaviors (including emotions, physiology, and deliberative or reactive actions) need to be identified. Here, it can be distinguished between encoding models that fall back on insights from cognitive science (e.g. models of how humans process emotions) and decoding models that specify an intended effect and are based e.g. on perception experiments but may also use cognitive models as “inspiration”. Whereas AI mostly deals with the first type, for graphics usually the latter is important. Since this report follows through all the way to rendering issues, this thereby includes procedural gesture and face animation, the simulation of tears, and rendering of wrinkles or skin tone changes, which will be discussed in sections 4 and 5.

Further, especially in mobile computing in combination with geolocation-based services, there is a recent trend in augmenting the real world with virtual information, which is made possible due to increasing processing power, bandwidth, and 3D-capabilities even on mobile devices. Thereby, fascinating new user experiences become possible, where e.g. a virtual character, as an augmented master teacher, explains and demonstrates the use of a new appliance. Hence, in case the agent shall be part of such a Mixed Reality application, besides the geometric registration between real and virtual objects as well as vision-based tracking techniques, suitable real-time rendering methods need to be incorporated too [GCHH04], that permit a seamless integration of the virtual character with the remaining real scene.

Since characters are an ongoing research topic, there al-

ready exist surveys on certain subquestions – the closest are outlined next. Real-time animation techniques in terms of their motion naturalness and the amount of control that can be exerted over this motion, with focus on animation systems in general without considering high-level control, are discussed in [vWvBE*09]. Emotions and emotion models, with focus on discussing the linguistic, socio-scientific, and psychological grounds are surveyed in [VGS*06], whereas rendering and implementation issues in general are only lightly touched upon while psycho-physiological reactions like blushing or crying are not mentioned at all. Approaches in automatic camera control are presented in [CO06], mentioning that the coordination of graphics and language poses a number of problems for camera control within an MDS – yet lenses, filters, and other visual effects important for expressing moods or directing attention are left aside. However, up to now no report has covered the problems and specificities of multimodal dialog systems (including the consideration of rendering methods to present the character in a visually plausible way or psycho-physiological effects like blushing or crying) nor presented all relevant research topics from a computer graphics point of view.

This report describes the aforementioned issues, particularly the main fields of research discussed in the following sections, which coarsely correspond to the numbers in the boxes shown in Figure 2. Considering all those disciplines, which generally are all research topics on their own, is a broad field. However, in dialog systems all these topics are connected with each others, but mostly only dialog and high-level behavior generation, interface languages, and character animation are considered here. Thus, our focus lies on going all the way “from models to rendering”. A main challenge is the connection between low-level graphics on the one hand and high-level behavior control on the other, since there is still a gap between behavior planning and concrete realization [KHGS10]. The dotted demarcation line moreover distinguishes between consciously controlled actions, namely voice and motor control, that in general are considered in an MDS, as well as unconsciously happening phenomena, that are not controlled via the central nervous system and usually are ignored in research, yet nevertheless important.

2. High-level Behavior Control

In this section we introduce the notion of multimodal dialog systems and then focus on the output part of such systems, more specifically on behavior planning and control and on emerging control languages.

2.1. Multimodal Dialog Systems

Multimodal dialog systems (MDS) extend traditional speech-based dialog systems with added modalities for input (pen, multitouch, camera, sensors, tangible artifacts) and output (2D/3D graphics, video, physical artifacts, robots)

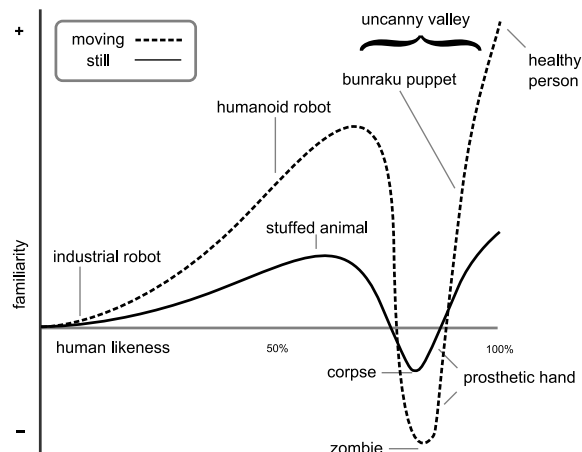


Figure 3: In the so-called uncanny valley the acceptance of anthropomorphic entities is worst (compare <http://www.androidscience.com/theuncannyvalley/proceedings2005/uncannyvalley.html>).

[KDD*05]. The goal is to enhance the interaction in various respects like robustness, ease-of-use, enjoyment and efficiency. Virtual characters can be considered as a natural “output device” for an MDS because they allow *symmetric* interaction, i.e. they allow “both the user and the system to combine the same spectrum of modalities” [Wah06, Wah03]. Virtual characters can and must utilize nonverbal communication like facial expressions, gaze, gestures and postures. This is both a potential and a challenge because the lack of natural behavior has the reverse effect of irritating users or even repelling them. This is often referred to as the “uncanny valley” effect, a hypothesis introduced by Masahiro Mori already in 1970 [Mor70]. His hypothesis states that as a figure is made more human-like in its appearance and motion, the emotional response from a human being will become increasingly positive and empathic, until a point is reached beyond which the response quickly becomes strongly repulsive. Figure 3 visualizes this relationship.

The question of whether virtual characters are beneficial in various specific domains (education, assistance for the elderly, sign language) is a research field in itself (cf. [Gul04, DvM00] for an overview). Moreover, the character’s behavior must not only be responsive and believable, but also *interpretable* [TMMK08], just like in face-to-face communication. However, it seems that adding a virtual character may be beneficial more in terms of motivation and enjoyment rather than improved task performance [MV02, LCK*97]. It is still under debate under which conditions motivation is increased without compromising the user’s task performance [MKK10].

For the control components of multimodal dialog systems this means a shift from natural language generation [RD00] to multimodal behavior generation [BKMW05], since the di-

alog management not only generates voice output but also the corresponding, fully synchronized nonverbal behavior [CBCV00]. In general, multimodal signals are characterized by their meaning and communicative function on the one hand as well as their visible behavior (e.g. shown through muscular contraction on a 3D facial model) on the other hand. For example, a deictic meaning (“here”, “there”, ...) maps to a deictic pointing gesture. Modalities are interrelated not only in meaning but also with respect to temporal organization (cf. [McN05, Ken04]). For instance, gestures are hypothesized to co-occur with the corresponding word or phrase in speech (the so-called lexical affiliate [Sch84]), although the question of exact timing is still under debate [Fer10]. Various techniques for synthesizing nonverbal human-like behaviors are reviewed in Sec. 3.1.1.

Albeit the architectural design of dialog systems as well as accompanying topics like natural language processing, discourse planning and the like are beyond the scope of this report, a brief overview and definition of relevant terms is necessary to understand the research field, which focuses more on the presentation component that is responsible for the perceptible visual output. Here, an embodied conversational agent (ECA) is a virtual agent that often has an anthropomorphic stature [CSPC00, PI04], whose cognitive and expressive capabilities simulate human capabilities, and which is capable of (interactively and usually autonomously) communicating with a user through verbal and nonverbal means [DCPPS02, KKM*06a].

Dealing with ECAs requires multidisciplinary collaboration between different fields of research like AI, linguistics, cognitive and social science, psychology, computer graphics, and so forth. The main issues from a CG point of view are character modeling, realistic real-time rendering, dynamic simulations, and the natural animation of face and body. One focus of this article is to clarify the connecting steps between low-level graphics on the one hand and high-level behavior control on the other. The next section will begin with the highest level: behavior control.

2.2. Behavior Control

A character’s behavior contains information about the content and expressivity of the communicative act, and it is not only determined by the communicative intention but also by the character’s underlying general behavior tendency. Such behavior generally is modeled following top-down approaches like the aforementioned goal-oriented application type. Since nonverbal communication as part of the human behavior always takes place, why “one cannot not communicate” [WBJ69], and thereby is an essential aspect of communicative acts, modeling of communicative behavior as such must be handled beforehand on a higher level and is not part the main part of this work, though the visualization component must be able to display this behavior in a flexible way. Thus, the need for higher level interfaces that allow a more

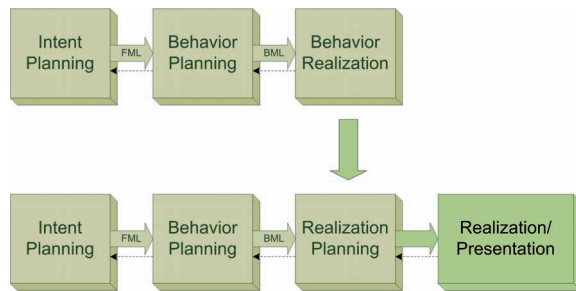


Figure 4: Stages of behavior generation according to SAIBA (top) and splitting of behavior realization into realization planning and presentation as proposed in [KHGS10].

abstract definition and control of object behavior call for better mechanisms of animation control and scheduling.

Therefore, within the SAIBA framework [VCC*07, KKM*06b] for interactive agents, three main stages of behavior generation were identified, that are mostly independent of the concrete realization of a character, namely intent planning, behavior planning, and behavior realization (see Figure 4, top row). This aims at replacing the previous monolithic or in-house architectures, as for instance used in [JK03], with a service-oriented software architecture that enables unified and abstract interfaces [HK09].

When we focus on the intent planning, we can identify several main approaches that we review in this section. To define the problem: Based on the perceived input from human user and virtual world on the one hand, and on the agent's goals on the other hand, the intent planning module decides what action(s) should be executed next. Such actions include high-level behaviors like speaking a sentence or walking to a target location, whereas lower-level behaviors like producing a gesture or changing posture are in the responsibility of the behavior planner. Also, the intent planner does not perform the actual realization of the actions. Instead, we assume that the output actions can be formulated on an abstract level (e.g. using BML language [KKM*06a]), and executed by dedicated realization engines (e.g. speech synthesizer, character animation engine).

2.2.1. Scripting

Manually written scripts are the simplest way of expressing behavior for a virtual character. However, the author has to program reactions to all possible situations and provide sufficient variations in order to avoid predictive behavior. Pedagogical agents such as Cosmo, Herman the Bug or the presenter Jack are examples of manually scripted agents [LVTC97, NB97]. Perlin's Improv framework [PG96] attempts to alleviate the variety problem by organizing the scripts into layers and groups. A script may call another script from a group in a lower layer which is then either selected randomly or using if-then rules. However, when introducing more powerful constructs like conditions, loops or

threads, a scripting language can quickly turn into a fully fledged programming language requiring expert knowledge.

2.2.2. Planning

Automated planning avoids the need to enumerate all possible situations and reactions to them. A planner receives as an input the current state of the world, a description of actions that change the state of the world (provided by the programmer and represented as plan operators with preconditions and effects) and a goal (represented as a state in classical planning or as a task in Hierarchical Task Network planning). The output of the planner is a sequence of actions that would, if successfully executed, bring the system from the current state to the goal state. Hierarchical Task Network (HTN) planners extend the description of the classical planning problem in order to reduce the time complexity. The goal of the HTN planner is to produce a sequence of actions for a task. A task is either primitive (represented by a plan operator) or compound (consisting of subtasks and a method that prescribes its decomposition into subtasks). The planner uses the methods to decompose compound tasks into smaller and smaller pieces until it reaches primitive tasks. André et al. [AR01] applied hierarchical planning for generating control scripts in a series of agent application ranging from a simple non-interactive presentation agent (PPP Persona) to interactive performances by several agents and multiparty scenarios [AR00]. In the latter case, a distributed planning approach was used where the performance was rather represented by the roles and the individual goals of the single characters as opposed to a hierarchically structured script. A similar approach was taken in [HR06, LPR07].

A significant amount of research has been devoted to real-time planning approaches that are able to cope with dynamically evolving situations within a narrative environment (cf. [RSY03, ADP06, CCM02]). For instance, Cavazza et al. [CCM02] combined top down planning approaches with reactive agent behaviors for coping with unexpected events of narrative relevance. Riedl et al. [RSY03] used a partial-order planner to account for unexpected user reactions that might require a re-organization of the narrative. Pedagogical agent Steve [RJ99], that features locomotion, object manipulation and tutoring capabilities, also produces its actions by a hierarchical partial-order planning. The Facade system [MS03] for interactive drama uses *reactive planning* with a Java-like language called ABL that allows the coordination of behaviors of several characters. *Anytime planning* is a recent, alternative method that was applied for virtual bots for computer games in the Excalibur project [Nar00]. It aims at improving reactivity and adaptation to unexpected situations by always providing a first immediate plan which is then improved if more time is available.

Planning approaches bear the benefit that they may be combined with an affective appraisal system. First, emotions can arise in response to a deliberative planning process (when relevant risks are noticed, progress assessed, and

success detected). For example, several approaches derive an emotion's intensity from the importance of a goal and its probability of achievement, see [GM04]. Second, emotions can influence decision-making by allocating cognitive resources to specific goals or threats. Plan-based approaches support the implementation of decision and action selection mechanisms that are guided by an agent's emotional state. Examples of embodied agents that integrate AI planning with an affective appraisal system include the Mission Rehearsal Exercise (MRE) system [GM04], the Hamlet component of [DCPPdR01], and FearNot! [ADP06].

2.2.3. Rule-Based Systems

In a rule-based system (RBS) the behavior of an agent is encoded by a set of condition-action rules, the *rule base*. An *inference engine* cyclically examines the conditions of all rules and selects a subset of those rules whose conditions are satisfied based on the working memory. One of those rules is then executed which may modify the working memory and thus trigger another rule in the next cycle until so-called quiescence is reached. Example systems are CLIPS <http://clipsrules.sourceforge.net> (C Language Integrated Production System) and JESS <http://herzberg.ca.sandia.gov> (Java Expert System Shell). Sport commentary agent ERIC [SK08] uses JESS for his reasoning, dialogue generation and affect appraisal. The real estate conversational agent REA [CBC*00] implements its deliberative module in CLIPS. The RoboCup commentator systems Byrne and MIKE [ABTI*00] partly employ RBS for generating emotional state and reasoning about events. Rule-based systems are well-suited in scenarios where knowledge processing is involved (e.g. building higher-level facts from low-level information). However, as rule bases grow, the procedural aspects of a RBS become very hard to predict, so that a hybrid approach may be more suitable (Section 2.2.7).

2.2.4. State-Based Systems

In this approach, the character's mind is both represented and visualized by states and transitions. Actions, attached to either a state or a transition, are executed as the graph is traversed. This approach has already been successfully used in the CSLU Toolkit for speech-based interaction [McT99]. In the CrossTalk and COHIBIT systems, interactive embodied agents are controlled by the so-called *sceneflow*, an extended hierarchical finite state machine (FSM) where a node represents an atomic state or a supernode containing another FSM [GKKR03]. Transitions can be conditional, probabilistic or interrupting (for exiting a supernode). Both nodes and edges may have pre-scripted scenes attached that specify dialogue and nonverbal actions. This approach bears resemblance to David Harel's *statecharts* [Har87]. Parallel Transition Networks (PaT-Nets) are another similar concept that incorporates facilities for parallelity. They were used in the area of character animation for the combined control of high-level behavior and low-level animation [BWB*95].

2.2.5. Connectionist Approaches

Several scientists experimented with biologically motivated methods that use an interconnected network of simple units. Percepts from the environment are fed into the input units and propagated through the network to the output layer in which the most active unit represents a decision. E.g., the Norms in the Creatures computer game [GCM97] are controlled by two *neural networks*, one for decision making (e.g. selects a command activate) and attention (selects an object to be activated), and one for selecting sensory-motor commands. In another system an autonomous virtual human is controlled by Tyrell's *free-flow hierarchies*, an ethologically motivated network architecture [dST05].

2.2.6. Multi-Agent Systems Architectures

Multi-agent systems research suggests several concrete architectures for controlling intelligent agents. BDI (Belief-Desire-Intention) is a cognitively motivated architecture where beliefs represent information about the world, desires are options available to the character and intentions denote goals that an agent is committed to. A planner is usually used to generate a sequence of actions based on the current set of intentions and beliefs. BDI was employed in agent scenarios modeling autonomous life-like behaviors [CT00] and social interaction behaviors [GVT00]. Brooks' *subsumption architecture* represents a purely reactive model [Bro91]. Complex behavior is decomposed into simple condition-action behaviors and organized into layers. If percepts from the environment satisfy conditions of several behaviors, the lowest one is selected.

2.2.7. Hybrid Approaches

Hybrid architectures combine several control methods to balance the needs for reactive and deliberative behavior. For instance, the REA agent [CBC*00] processes inputs either by a set of hardwired reactions which result into an immediate output (e.g. agent's gaze tracks user's movement) or by a deliberative planning-based module (e.g. selection of utterances according to a communicative goal). The CrossTalk system [KKGR03] compiles a hierarchical FSM into plan operators which can be used in a classical plan-based approach at runtime, whereas the ITeach system runs an FSM and a RBS in parallel, synchronizing them using shared variables [MK09]. The MRE system [SHG*01], a virtual reality training environment for military operations, combines several control methods based on character type: agents with limited behaviors are pre-scripted, agents directly interacting are driven by a planner [RJ99]. Furthermore, the story is structured and controlled by a FSM similar to the sceneflow in SceneMaker [GKKR03]. Hybrid approaches are often necessary when scenarios become complex. Different aspects can then be handled by appropriate technology. For instance, knowledge processing aspects are best handled by a RBS, whereas procedural aspects are best modeled with a FSM or statechart (ideally with a graphical interface).

2.3. Control Languages

Different kinds of behavioral models depend on the level of autonomy of the character and on whether body and mind are considered independent or not [DCPPS02]. Control languages serve as reusable representation of agent behavior and separation between modules that implement different functions, for instance behavior planning and realization. The BEAT toolkit for the automated generation of nonverbal behavior used multiple languages to pass information from module to module [CVB01]. Out of this framework emerged the SAIBA model where the *functional markup language* (FML) [HKM*08] is used to encode the communicative intent without referring to physical realization and the *behavior markup language* (BML) specifies the verbal utterance and nonverbal behaviors like gesture, posture and facial expression [VCC*07, KKM*06b].

By defining an additional dictionary of behavior descriptions, the “Gesticon” [KP04], the language distinguishes between abstract behavior definitions and concrete realizations. MURML [KKW02] and APML [dCPPS04] are, like BML, specification languages for physical realization. MURML allows describing gestures by defining spatiotemporal constraints and submovements of a gesture stroke. An application example is demonstrated with the anthropomorphic agent Max in [JK03]. MPML/MPML3D [PSI04] was designed for web-based scenarios and codes verbal and nonverbal behavior, presentation flow and integration with external objects. VHML [Mar01] is an XML-based language which consists of several sub-languages for describing the character, like GML for its gestures, FAML for facial animation, BAML for body animation, EML for emotions, etc. Here, the Emotion Markup Language was designed to represent the emotional states to be simulated by a user interface or of a human user in a standardized way. Examples for describing emotion-related behavior with EmotionML are given on <http://www.w3.org/TR/emotionml/\#s5.1.3>.

Since languages like BML employ concepts like relative timing and lexicalized behaviors, [HK10] outlined the need for an additional declarative *animation layer*, a thin wrapper around the animation engine and situated below higher-level behavior control layers for abstracting away from implementation details while giving access to the functionality of the engine (cp. Figure 4). Exemplarily their proposed system architecture is shown in Figure 5.

For developing interactive virtual humans on the graphics side not only the geometric model and some basic ways of animating it have to be taken into account, but also aspects belonging to different levels of abstraction. Thus, in [IC05] the authors propose a generic, layered software architecture that allows focusing on the behavioral aspects, whilst providing animation models that also include collision detection and path planning. In [YPWP05] a VRML based system consisting of three layers for animating characters is described. Whereas the lowest layer controls the joints, the

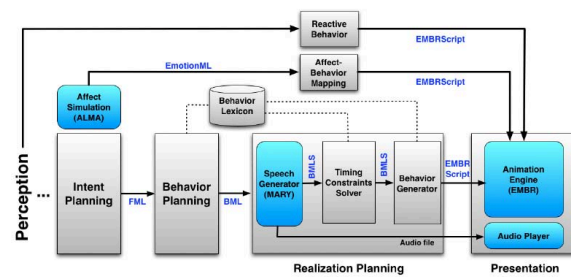


Figure 5: Overview of an exemplary system architecture (internal view) as proposed by [KHGS10]. ©2010 Springer.

middle layer combines a predefined schedule and different joint transformations to skills like “walk” or “open door”. The highest level was an English-like scripting language for expressing the composition of skills and for hiding the complexity of lower layers. A similar approach is proposed in [HEV03], although in this work the authors use their scripting language already for composing primitive motions based on operators like ‘repeat’, ‘choice’, ‘seq’ and ‘par’.

Likewise, [JB08, JB09b] proposed to further split the presentation component into a hierarchy that can be roughly categorized into a control layer for behavior description and animation scripting, and an execution layer for providing the low-level building blocks that are necessary to fulfill the requests of the control layer. Their framework furthermore builds on the open ISO standard X3D [Web08], which is used as the application description language. This work thereby follows [GRA*02] who outlined, that animation standards such as H-Anim [Web05] (which specifies the structure and manipulation of articulated, human-like characters) facilitate the modular separation of animation from behavioral controllers and enable the development of higher-level extensions. Yet, in [GRA*02] it was also remarked that the main problem of H-Anim is the lack of a general behavior and animation control API in its corresponding X3D language binding, which is tackled by this work [JB08, JB09b].

Another approach is to use a scripting language like Python to give access to animation functionality as done in PIAVCA [GS10]. Similarly, [Tüm07] utilized Lua, a lightweight and embeddable scripting language that is often used in games, for scripting his animation engine. However, the output of high-level dialog engines are descriptive directions rather than algorithmic procedures, which makes imperative languages like Python and Lua inappropriate as output of the planning stage. A comparison of common markup languages e.g. can be found in [OOdPC*09].

3. Real-time Character Animation

Virtual characters in general and character animations in particular are obviously an important and ongoing research topic. For example, in their survey, di Giacomo et al.

[GMMT*07] discussed techniques to automatically generate a skeleton from a character representation for animating it with bone-based animation techniques. To fulfill the requirements of higher level control layers, a flexible animation system is necessary. In this context, van Welbergen et al. recently surveyed real-time animation techniques in terms of their motion naturalness and the amount of control that can be exerted over this motion [vWvBE*09], though this report focused on animation systems in general without considering high-level control.

3.1. Body and Facial Animation

Model-free approaches [SH07] are usually restricted to facial animation, whereas body animations mostly follow model-based approaches that expect a predefined structure. Therefore, the character model first has to be rigged, i.e. a hierarchical set of bones, which is used for animations only, needs to be associated with the characters surface representation, namely the skin [GMMT*07, GS10]. For real-time body animation, mostly the well-known skins and bones approach is used. In addition, basically two types of animations can be distinguished, namely data-driven models and procedural methods [GS10].

Facial animation usually is done with Morph Targets or Blend Shapes respectively [ABM00], whereas in the MPEG-4 standard facial expressions are modeled by modifying certain feature points [PP04]. To overcome the uncanny valley effect, [ARL*10] presented a photorealistic digital character that was animated through video performance capture. Therefore, shape and reflectance of a real face was digitized in more than thirty poses showing different emotions, gaze directions, and visemes, including skin appearance and wrinkles by using a light stage [DHT*00]. However, only parts of the final scene were virtual, and setup, animation, and rendering required several months and artists.

3.1.1. Nonverbal Behavior Synthesis

Research on nonverbal behavior synthesis (gesture, posture change, gaze and facial expression) tries to maximize two opposing goals: naturalness of motion and flexibility in motion control [vWvBE*09]. The two concerns are complementary and no proposed motion generation system has yet been able to generate convincing motion while offering a high level of control and flexibility in the design or specification of motion.

Data-driven approaches like motion graphs (Sec. 3.1.2) rely on key-frame data, where a motion sequence can be based on motion capture data or is defined for specific key-frames by a skilled animator. They achieve a high degree of believability since MoCap data preserves natural movements, but is expensive, requires the use of special suits, and is limited to capturing the skeletal motion of the human body, leaving the dynamics of cloth and hair aside [SH07].

Furthermore, its main drawback is inflexibility and retargeting animation data [Gle98] is often critical. Moreover, data-driven animations cannot easily be modified: slight changes in the performance requirements imply a new and costly motion capture session.

One possibility to add control is to extract style components from existing motion clips (e.g. angry, weary) using dimension reduction techniques and to re-apply these components on other clips to endow them with the qualitative aspects of the initial motion [HG07]. With “action capture”, in [AHJV08] the problem of goal-directed motions is tackled at a higher level, in that a certain animation is considered as a skill including the interaction with scene objects. However, motion capture approaches are quite successful when the input is also a low-level signal. Levine et al. use purely acoustic features of the speech signal to synthesize appropriate gesture motions on-the-fly [LTK09, LKTK10].

Knowledge-driven approaches use procedural animation where the human performance is generated from scratch in order to satisfy basic constraints like hand positions, head orientation or foot steps. In such cases, the specification of the motion involves a set of spatial constraints in time. In its more basic form, motion generation algorithms only use analytic [TGB00] or iterative algorithms [Wel93] derived from the control of robotic articulated structures. Such methods can lead to fast and robust animations which can be applied in games and other interactive applications, though the high number of degrees of freedom is an issue. However, the generated animations suffer a lack of naturalness which makes the user uncomfortable in a user-agent interaction scenario.

Improvements of procedural animations take into account theories of human motion [CCZB00], psycho-sociologic studies [HMP06], affective state [LM06] or manually analyzed video material [NKAS08, KNKA07]. Other approaches focus on physical principles of articulated structures. For instance, a humanoid figure is more convincing when it tries to keep its balance while performing an action [HWBF95]. Studies dedicated to physical simulation took into account musculo-skeletal properties in order to mimic human aspects of motion like the stiffness of a gesture [NF02]. This work was further developed into a framework that allowed a range of movement properties to be sketched for a character and automatically applied to its motion [NF05]. A recent trend in nonverbal behavior synthesis is to design behavior inspired by the literature, and then validate the perception effect with user studies. This has been successfully applied to gaze behavior [CGV09, KG08, NBA09] and gesture [NWA10].

3.1.2. Motion Planning and Synthesis

In [Rey99] motion behavior is divided up into three hierarchical levels – action selection through higher level goals, steering, and locomotion. This work focuses on the second level, path determination for autonomous agents (or non-

player characters/ NPCs as they are called in games), by describing a set of steering behaviors such as “seek”, “path following” and “obstacle avoidance”, leaving animation aside. In contrast to motion planning that only refers to high-level goals, motion synthesis denotes the generation of the low-level details of a movement [PKL08].

To achieve flexible high-level character control with natural movements, in [SKG05] therefore a multi-level (yet off-line) approach was presented. First, at the highest level, path planning takes place, e.g. by simply using A^* or, as proposed in [SKG05], by utilizing probabilistic roadmaps (PRM). These randomly sample the so-called “configuration space” \mathcal{C} for finding possible configurations (i.e., poses of an articulated system) that make up the roadmap (cf. [CLH*05, p. 202 ff.]). At the second level, the resulting path is approximated by a composition of motion clips obtained from searching a motion graph, which requires a preprocessing step that annotates all clips for finding possible transitions to build the graph. The third and lowest level deals with adjusting the motions, e.g. by blending motions together, to follow the path at the best.

Motion graphs [KGP02, LCR*02] are directed graphs where all nodes correspond to motion fragments and can be seen as generalization of blend trees (cp. [Eds03]). The goal is to obtain realistic human motion including the subtle details of human movement that are not present in procedurally generated motions. Although these methods lead to convincing results even for on-line motion generation, they still require preprocessing, are computationally expensive and high memory consumption is an issue [MP07]. Thus, most approaches are still targeting at off-line motion generation like the one presented in [SH09]. Complementary to motion graphs are parameterizable motions [RCB98, SCFRC01], where the focus lies on generating parameterizations of example motions, such as walking, jogging, and running [PSKS04]. In general, they also tackle the problem of design freedom, since customizing given animations per se is hardly possible [vWvBE*09].

3.2. Overview of Virtual Character Systems

In the past two decades, several research groups developed reusable virtual character systems that can be employed in multimodal dialog systems. Greta started out as an MPEG-4 based facial animation system but has been extended to full-body motion, complying to the BML control language [Pel05]. SmartBody [TMMK08] was the first BML animation engine and has been used in projects like MRE. EMBR [HK10, HK09] is a more recent BML-compliant player that offers an animation control language to have more fine-grained control, e.g. for highly reactive motions like gaze following, and has recently been used for sign language synthesis. Max [KKW02, JK03] is an animation engine that was controlled by a language called MURML which allows a highly flexible definition of gesture form and timing.

MARC [CBM09] is an MPEG-4 [PF02] based facial animation engine that also models facial wrinkles and is being extended to full body motion. Elckerlyc [vWRRZ10] is a BML compliant animation engine that uses physical simulation to enhance motion synthesis and extends BML for reacting to real-world events. Piavca [GS10] is a character engine that targets integration in virtual environments and can be controlled with Python. Similarly, the system described in [JB09b] is embedded into the Instant Reality framework [IR10] and also targets at the integration with virtual environments, though here the focus is more on rendering aspects. As described in [JB08] it can be controlled with a declarative language called PML [GSI*04, KG07, JK07].

The main focus of multimodal dialog systems are output modalities and thus lies on gestures, facial expression, and speech, yet rendering and psycho-physiological processes are not yet covered by all engines. An overview of the more technical aspects of current character engines, including a comparison concerning skeleton setups, animation generators, and control, was recently given in [GS10]. Based on the results of Thalmann et al. and for integrating further research the VHD++ framework was developed [PPM*03]. It is a service-based middleware solution and thereby extensible on the code level, but many features like cloth and hair simulation are not public and moreover the toolkit is not suited for non-graphics people.

Likewise, Egges’ research [Egg06] is based on VHD++. Here, the focus lies on the development of an animation model for interactive dialog applications, paying particular consideration to facial expression, gestures, and idle motions. The model combines several approaches, like motion captured animations with procedurally generated ones, while also considering emotional states. The framework and the underlying motion-synthesis-from-analysis technique are also described in [MTE06]. Tümmler [Tüm07] stated, that in the area of virtual characters a lot of good but isolated applications exist, which normally can be hardly combined to a total solution and in practice often have to be re-implemented. A comparison of current animation toolkits can also be found in [Tüm07, p. 45]. But all of them utilize proprietary formats, define their own content pipelines, etc.

4. Dynamics of Human Hair

Flexible animation systems require simulating resultant dependencies like hair movements [ND05, WBK*07, YT10], e.g. when a character moves his head, or “nervously” runs his fingers through his hair. Furthermore, MoCap is limited to skeletal motions, leaving the dynamics of cloth and hair aside, whereas the latter can be an integral part of a gesture.

4.1. Modeling and Simulation

To create convincing human hair there are basically four problems to solve: modeling and styling, hair dynamics,

collision detection and response, and finally hair rendering [NTSP02, WBK*07]. Presently a seamless transition between these categories is problematic, since the fewest systems are self contained and differ in their geometrical representations, animation methods, and lighting models. Thalmann et al. [NTSP02] classify hair models into several categories. The first one contains explicit models, where each hair strand is considered individually [iAUK92, DMTKT93]. However, this type is problematic, since a human being usually has around 100,000 hairs.

The next category are cluster models, which utilize the fact that neighboring hairs have similar properties and tend to group together. They can be further divided up into hierarchical and flat models. A hierarchical level of detail representation is proposed in [WL03, WLL*03], in which the hairs are either represented as precomputed strips, clusters, or strands, depending on visibility, distance to the viewer, and velocity (see Figure 6, right). More common are non-hierarchical schemes in which clusters are represented by generalized cylinders [KN02], prisms, and polygon strips. Particle systems [BCN03] can be seen as an extension to clusters. The last category regards hair as a continuum [HMT01].

To simulate complex hair styles in real-time, external environmental forces like wind or gravity must be considered, too. In [VMT04] it is proposed to use a free-form deformation (FFD) grid that contains the hairs and which during simulation is being deformed using a mass-spring-system. Bertails et al. [BAC*06] presented a physically-based model, where each hair is represented by a so-called super-helix. This also allows simulating curly hair but is not real-time capable. In [Yu01] curliness is added by modulating the strands with offset functions. In the model of [KHS04], a hair patch is composed of quadrilateral segments, and curly effects are achieved by projecting the vertices onto an imaginary cylinder surrounding the strand.

A method based on differential equations is the modified cantilever beam simulation originally proposed in [iAUK92] for computing the hair bending of smooth hairstyles during the modeling process. In their model, a hair strand is modeled as an open, serial kinematic multi-body chain with anchored root and segments that are connected through joints. As visualized in Figure 6 (left) the strand then is deformed by obtaining the new angular positions (Θ, Φ) of each control point. A robust and easily parameterizable simulation method was proposed by [JRKL05, JK05]. Neighboring hairs are combined into wisps and animated with a simplified cantilever-beam-based simulation system that acts on the kinematic chain defined by the skeleton hairs and which runs numerically stable and with real-time update rates.

Tariq [TB08] presented an impressive real-time approach for simulating smooth hair styles, which runs completely on the GPU and exploits modern features such as stream output to optimize performance. The latest DirectX 11 features

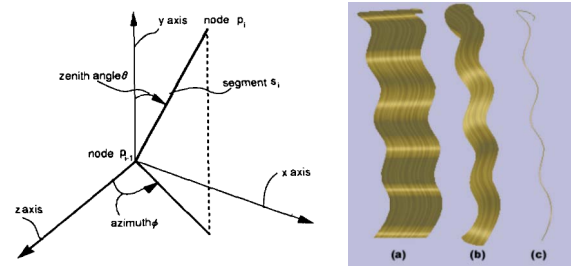


Figure 6: Left: polar coordinate system for modeling/animating a hair segment (taken from [iAUK92]). ©1992 ACM. Right: LOD representation of hair [WLL*03]: a) Strip, b) Cluster, c) Strand. ©2003 IEEE.

(or Shader Model 5.0/ OpenGL 4.1 respectively) further allow to directly tessellate the base hairs on the GPU [YT10], while simultaneously handling collisions for multi-strand interpolation. But currently this method requires expensive high-end graphics hardware. Absolutely inevitable is the treatment of hair-head collision. Whilst geometry traversal, hierarchical or grid based schemes, and vector fields offer more precision, for real-time applications a head can be approximated sufficiently with the help of spheres or ellipsoids [iAUK92, ND05]. Hair-hair collisions for interactive applications are still mostly ignored or quite coarsely approximated, e.g. by bounding volume hierarchies [WL03]. Yuksel and Tariq [YT10] handle inter-hair collisions by voxelizing the hair volume, and by then pushing vertices falling in high density areas into the direction of the negative gradient of the voxelized density field.

4.2. Hair Rendering

Rendering covers the full range from drawing polylines and alpha-textured polygons [Tüm07] over heuristic local lighting models for anisotropic materials [Sch04] up to physically and physiologically correct illumination solutions [MJC*03]. Self-shadowing can be achieved with the help of an opacity map [KN01], which discretely approximates the intensity attenuation function for encoding a fragment's opacity value, mostly in a preprocessing step [KHS04]. An improvement where the opacity layers were adapted to the form of the hair volume recently was proposed by Yuksel and Keyser [YK08]. An often referred reflectance model for dark hair, which exhibits higher reflectance than transmission and almost no self-shadowing, can be found in [KK89].

Based on Marschner et al.'s [MJC*03] results, Scheuermann [Sch04] modified this model for the use in shaders by perturbing the hair tangent for shifting both highlights. To overcome this more phenomenological approach, where multiple scattering is faked with an ad-hoc diffuse component coupled with transparent shadows, Zinke et al. [ZYWK08] recently presented the concept of dual scatter-



Figure 7: Real-time animation of weeping as another example of affective behavior. Note how the extreme closeup is used to show the characters' emotions in greater detail.

ing, which splits the multiple scattering computation into a global and a local multiple scattering component.

Rendering hair as polylines or curves (e.g. [ND05]) has several drawbacks like aliasing. In addition, antialiasing of line drawing and alpha blending in general needs correct visibility ordering. In [Sch04] this is solved by drawing the pre-sorted hair patches in several passes. Kim and Neumann [KN02] suggest an ordering algorithm in which each line segment belongs to a slab perpendicular to the viewing vector. A comprehensive overview on related techniques is also given in [WBK*07].

5. Emotion Visualization

Emotional state and discursive elements are communicated through gaze and facial expressions, and the human eye is extremely familiar with them. Hence, besides the advances in MoCap, which can be accurate enough to capture even slight movements, new rendering techniques to ensure realistic skin models need to be considered, too. Another issue only barely considered here is affective aspects. Temporal variations of facial coloration due to blushing and paling, along with sweat and tears are important for simulating consistent, believable expressions [KMT94].

5.1. Psycho-physiological Factors

A rather elaborate discussion on emotions and emotion models in the context of simulating the expressions of virtual characters was given by Vinayagamorthy et al. [VGS*06]. But the focus of this work lies on discussing the linguistic, socio-scientific, and psychological grounds, whereas rendering and implementation issues in general are only lightly touched upon while psycho-physiological reactions like blushing or crying are not mentioned at all. However, faces do not only have static features like skin color and feature size, or slowly changing signals like wrinkles, but they also exhibit rapidly varying signals that are often caused by emotions like joy or sadness [KMT94]. This also includes the simulation of tears, rendering of wrinkles, and skin tone changes (compare Figure 7).

5.1.1. Facial Expressions and Emotion Theories

People not only are influenced in their thinking and actions by emotions but also are attuned to recognizing human

emotion [Sch81, Ham06]. Thus, virtual characters that display emotions are critical for believability, because emotions create the difference between robotic and lifelike behavior [VGS*06]. Emotions are a physical and psychological reaction to external and internal factors. They are reflected in the facial expressions, gestures, voice, diction, and a person's behavior in general, and usually cannot be controlled consciously. There exist various emotional models. The anatomically oriented Facial Action Coding System (FACS, [EF78]) distinguishes the following six emotional expression groups in conjunction with their corresponding geometric "deformations": surprise, anger, sadness, disgust, pleasure, and fear.

Plutchik [Plu80] developed a psycho-evolutionary theory showing eight primary human emotions and extended Ekman's model by adding two other emotions: acceptance and anticipation. Whereas discrete models are frequently used for displaying emotions, dimensional representations are often used for emotion recognition. The OCC theory [OCC88] belongs to the class of appraisal theories of emotions, which here can only be caused cognitively through a subjective and continuous evaluation of the environment. Because emotions including different intensities can be classified based on rules and decision trees, the OCC model is commonly used in AI for triggering emotions [BAW10].

5.1.2. Psychological Foundations

Whereas emotional variations of gestures and facial expression have been subject of extensive research, a more unattended field are psycho-physiological processes like the change in skin color, which can occur when an emotion is very strong. In this context basically two types can be distinguished. Blushing [KMT94, Mar96] is an uncontrollable reflex, which usually occurs in a social context, when a person feels ashamed or embarrassed. An average blushing takes $\Delta t = 35$ seconds. After 15 seconds, it has the strongest intensity and then it decreases again. However, blushing is not just an expression of emotions, but also occurs during physical effort. Thus, it is further distinguished between blushing (due to psychological reasons) and flushing (due to physiological reasons). Here, regions with many blood vessels like cheeks, ears and forehead, get more reddish. In a study of Shearn et al. [SBH*90] it was found that there is a correlation between blushing and increase of temperature in the cheek region – however, still no physiological or computational model exists.

The same goes for pallor [KMT94]. In contrast to the central nervous system (CNS), which is responsible for the conscious control of motor functions, the autonomic nervous system controls unconscious inner functions that can result in physical reactions like blushing and similar phenomena, which sometimes are described by the term vascular expressions. Other vegetative functions are sweat or weeping. Here, adults usually cry due to certain events or moods such as grief, anger, and joy. But this depends on culture, gen-

der, personality, physical state, hormones, social factors, and may also serve a communicative function [VC01].

5.1.3. Emotions in Computer Science

Human emotions are an important element in a communicative situation and thus should also be modeled to achieve plausible virtual characters with consistent behavior (see section 5.1.1). Here, the more manlike a virtual character gets, the more people expect emotional behavior. Therefore, models that can be used for the automatic synthesis of realistic communicative behaviors need to be considered. These include deliberative and reactive behaviors as well as physiology simulation and emotions. A comprehensive overview on computational models of emotion is given in [MGP].

Here, it can be distinguished between encoding models that fall back on insights from cognitive science, such as models of how humans process emotions, and decoding models that specify an intended effect. The latter may also use cognitive models, but they are mainly based on perception experiments etc. Whereas AI mostly deals with the first type, for graphics usually the decoding models are important. For instance, in the study of Buisine et al. [BWG10], it was shown that concerning decoding performance the most effective combination of speech and emotional expression is to temporally position the facial expression at the end of an utterance, while visual realism is perceived higher when the expression is shown during speech instead.

Gestures and facial expression reflect emotional behavior. But whereas both have been subject of extensive research, a more unattended field however beyond standard mesh-based animations is the change in color etc., which can occur when an emotion is very strong. But usually only body functions controlled by the central nervous system (CNS) like voice and motor response are considered in computer graphics by generating or playing-back different body animations and facial expressions. The latter are well understood and categorized by psychological models like the well-known FACS (cp. section 5.1.1), though communication effectiveness is still an issue in virtual agents research [BWG10]. Albeit with the help of modern graphics hardware the more subtle changes concerning face coloring can be covered, too, currently not much attention is paid concerning this topic, although it was shown that correct coloring and texturing can enhance the perception of certain emotions [EB05].

Blushing and pallor can be achieved by e.g. blending color values with every vertex along with its position and normal. Therefore, in [Pat95] a system is proposed, where the facial coloration is realized by changing the vertex color according to its position in the local coordinate system of the head. The amount of colorization is controlled by the emotional state of the virtual character. Alternatively, similar to a bump map for simulating wrinkles a blush texture map can be used as originally proposed by Kalra and Magnenat-Thalmann [KMT94], where another facial animation system based on predefined



Figure 8: Texture stack for changing face color [JK06] (left) and combination of affective facial expression (distorted lips and eyebrows) with skin tone changes for simulating rage (right) compared to a neutral expression (middle) [KG07].

image masks defined by Bezier patches for creating texture maps is introduced.

Although being quite outdated concerning their rendering methods the prime contribution of these works was to point out that realistic skin rendering also requires considering changes of skin color dependent on physical conditions and emotional state. Since then, usually only meso-scale geometry deformations such as wrinkles or pores were considered, mostly in the context of aging processes (for instance [OPV*99]), but also concerning emotions, like in the MARC system presented by Courgeon et al. [CMJ08], which also includes an automatic wrinkles calculation. A real-time method for animating tears and changes in skin coloration based on pre-defined key-frames encoded in a 3D texture was presented in [JK06] (see Figure 8, left). The authors later proposed a classification model for visually distinguishable emotional states caused by vegetative functions [JWKF09] to ease high-level parameterization.

Adding subtle changes in the facial color that relate to mimic skin distortions can help improve realism, although, as was shown by MacDorman et al. [MGHK09], the more texture photorealism and polygon count increased, the more mismatches in texturing etc. resulted in making a character eerie. Thereby, in the study of Pan et al. [PGS08] on human participants' reactions towards blushing avatars, one of the findings was that people noticed the avatar's cheek blushing, due to shame, more than whole-face blushing. But as an even more important outcome, the results suggested that people were less tolerant if only the cheeks were blushing. Obviously the latter was not convincing as a blushing response, and thus this type of blushing was worse than having no blushing at all.

But nevertheless, the study indicated a strong correlation between whole-face blushing and sympathy. Although the participants noticed the whole-face blushing less, they felt increased co-presence with a whole-face blushing avatar even though they have not been consciously aware of it. This corresponds with the findings described in [DdJP09] that a blush can remediate others' judgments after clumsy behavior. Likewise, the experimental study on emotion perception recently presented by de Melo and Gratch [dMG09]

suggested that considering physiological manifestations are especially useful to convey intensity of emotions. Another experimental study recently was conducted by [JW10] and indicated that considering skin changes can help improving the perception of certain emotions. Especially the emotional class *Anger/Rage* had a significant detection improvement, probably due to the proverbial red head. Moreover, the addition of physiological characteristics in general improved the average perception of emotions, whereas especially the male participants showed an improvement of around 20% in their average recognition rate.

In AI simulating emotions is an important topic, and due to its computability simulation is often based on the appraisal-based emotion model presented with the OCC theory [OCC88]. Thus, with ALMA Gebhard [Geb05] introduced a layered model of affect for enhancing simulated dialogs based on the OCC model, whereas emotions are calculated within the three-dimensional, continuous PAD space (i.e. pleasure, arousal, and dominance). In this regard, affect influences the character's mind. Based on the particular moods and emotions, dialog strategies and communicative behaviors are chosen. In his model three types of affect based on different temporal characteristics are distinguished: first emotions, to which facial expressions belong and which are short-term affects; then moods; and finally personality, which specifies a character's general behavior.

As was explained in Klesen and Gebhard [KG07], the affective state is then used to control gestures and facial expression or even facial complexions in real-time (the latter being realized following [JK06]). Whereas in [KG07] emotions are used to control facial expressions, skin tone changes, and other affective animations like weeping, moods are mainly reflected by postures. Therefore, these are used to control the posture and idle behavior (e.g. breath or eye blink rate). Exuberant characters for instance show more body and head movements than bored ones, who might only look at their virtual watch from time to time.

In their WASABI affect simulation architecture, Becker and Wachsmuth [BAW10] further differentiate between primary and secondary emotions in order to account for cognitively more elaborated emotions. Recently, [GAP*10] presented a graphical representation of human emotion that is extracted from text sentences using data mining statistic. Other relevant models, including the influence of personality traits and social circumstances, are also discussed in [VGS*06]. Also in robotics, emotional aspects like facial expression, and very recently even vascular expressions, are considered, which may indicate that this topic will be developed further in the future.

5.2. Rendering Issues

To plausibly present such psycho-physiological factors, we'll also have to cover some rendering issues here.

5.2.1. Skin Rendering

Skin is a multi-layered non-homogeneous medium with translucent layers that have subsurface scattering properties [INN05]. The colors of facial skin are mainly due to the underlying anatomic structures: muscles, veins, and fat all are visible through the skin's translucent layers. Besides greasy secretion and moisture the top-most skin layer is also characterized by the fine vellus hairs that are responsible for the "asperity scattering" [KP03]. But most of the light travels into deeper layers, which gives skin its soft appearance, why subsurface scattering is one of the most important phenomena of human skin [INN05]. A computational model for subsurface scattering as a simplification of the general volume rendering integral was presented in [HK93]. To account for both, Mie and Rayleigh scattering, in their model the material phase function for representing the directional scattering of incident light is described with the empirical Henyey-Greenstein function [HG41].

Based on an offline skin rendering approach [BL03], which utilizes the fact that for diffuse reflectance the outgoing radiance is spatially blurred, in [SGM04] a technique is proposed that approximates subsurface scattering on a more phenomenological level by blurring the diffuse illumination in texture space. To further emphasize light bleeding in case of backlighting, so-called rim lighting is applied by additionally adding the dot product between light and view vector that is scaled by a Fresnel term. Yet another approach is the extension of the commonly used BRDF by measured skin data (e.g. [KB04]) and a real subsurface scattering part. In [JMLH01] it is proposed to split subsurface scattering into a single and multiple scattering part. The complete BSSRDF then is a sum of both scattering terms. With limitations, mainly by only taking local scattering into account and evaluating the equation in image space, the model can be used for real-time environments [MKB*03].

Another method, more practical for real-time application, is based on maps recording the distance of a point seen from the light source. This depth map is used to measure the covered distance of the light ray inside a given geometry, and can be regarded as an approximation of path tracing. Such a technique of approximating extinction and in-scattering effects is presented by Green [Gre04] and is based on an offline approach for Pixar's RenderMan system described in [Her03]. Additionally an attenuation look-up texture can be used that maps the normalized light distance to a color ramp. Besides the limitation that only homogeneous materials are considered, only the first object in sight is recorded to the depth map, which can cause artifacts.

A good overview on real-time, GPU-based skin rendering techniques can also be found in d'Eon and Luebke [dL07]. For simulating scattering, the authors further present a method that combines the aforementioned texture space diffusion approach [BL03, SGM04] with modified translucent shadow maps [DS03] (similar to Green's

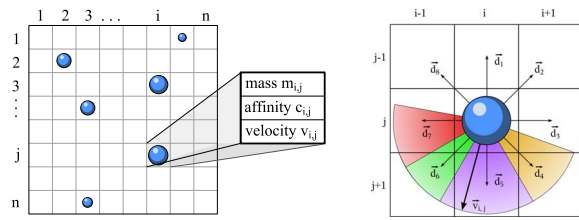


Figure 9: Physical properties of water droplet for every grid element (left) and possible directions d_k to neighboring grid elements and velocity vector $v_{i,j}$ (right), after [KIY99].

method [Gre04]). Here, texture space diffusion is done with the help of sum-of-Gaussians diffusion profiles for a three-layer skin model. In this regard, a diffusion profile $R(r)$ describes for all color channels, how the incident light attenuates for each distance r around a surface point. For rendering, each diffusion profile is first approximated by a sum of six Gaussian functions as pre-process. Then, the diffuse illumination is convolved (or blurred) in texture space for each Gaussian kernel. After that, the blurred textures are combined according to $R(r)$.

Likewise, Jimenez et al. [JWSG10] recently proposed a scalable approximation method for subsurface scattering, which is intended for game contexts and produces translucency effects at high frame rates. The main difference to the previous approach is the idea to simulate scattering in screen space as a post-process instead in texture space. Exploiting the observation that the effects of subsurface scattering are more noticeable on small curved objects than on flat ones, in [KDM10] subsurface scattering is simulated using a curvature-dependent reflectance function and a local illumination model. Though lighting is evaluated on a phenomenological level, less pre-processing than in [dL07] is required and rendering is only around 5% slower than Blinn-Phong as no multipass is needed.

5.2.2. Flow Behavior of Blood, Sweat, and Tears

The appearance and flow behavior of droplet flows is different from typical fluid dynamics and mainly influenced by factors such as surface adhesion or the formation of minimal surfaces. For interactive applications the animation of droplet flows can be mainly based on phenomenological observations. Thus, Kaneda et al. [KKY93] mentioned the following factors as being most relevant here:

Flow condition: A drop begins to flow down a slope, if it exceeds a critical mass, based on the amount and viscosity of the fluid as well as the angle of inclination of the surface.

Meandering flow: When flowing down, the drop makes meandering motions, which seem to be random, but are caused by taking the line of the least resistance.

Water trails: Drops leave a trail of liquid consisting of smaller drops behind, and because of the fluid loss, the drops finally grind to a halt. The resulting path then impacts the be-

havior of other drops, because here the resistance is less than at dry places.

Merging drops: Confluent drops merge to one, whose velocity depends on the velocity and mass of the original drops.

Based on these factors, they developed an empirical model for simulating droplet flow [KKY93, KIY99], by covering a surface with a discrete 2D grid. In every grid cell, drops can originate, jump to the next grid cell, and merge with other drops (compare Figure 9). First, the drops are placed on the grid (maximal one water droplet per cell), and then, drops and grid elements are initialized with their physical properties (mass, affinity, and velocity). The water affinity denotes the surface roughness and is used for meandering. After that, the most probable next flow direction d , based on Newton's second law of motion, the given water affinity, and the already present wetness, is evaluated stochastically. Based on Kaneda's model, in [Tat06] a GPU-based implementation for simulating rain-drops on window panes was presented.

A related method for animating the flow of water droplets on structured but flat surfaces was proposed in [JH02]. Here, the drops are represented by 3D particles and affected by the underlying bump mapped surface. Instead of real geometry, for speed-up the bump normals are used to control the motion of the droplets. A real-time method for animating tears and sweat based on pre-defined key-frames encoded in a 3D texture was presented in [JK06]. Recently, another algorithm to simulate tears based on Smoothed Particle Hydrodynamics, which acts on a 3D grid and therefore is not restricted to image space, was proposed by van Tol and Egges [vTE09a, vTE09b]. The simulated fluid then is visualized by generating a mesh each time step, using the marching cubes algorithm [LC87] for determining the resulting isosurface. But as opposed to image-space-based approaches (e.g. Tatarchuk [Tat06]), it is computationally more complex and does not scale well with the number of droplets.

An approach that was directly intended for GPU implementation, recently was proposed by El Hajjar et al. [HJGP08]. Similar to Kaneda et al. [KIY99], the authors also use a grid for describing the fluid distribution on the surface. But here, a grid element does not denote a single drop (as shown in Figure 9), but a much smaller amount of fluid that, together with neighboring elements, makes up a complete drop. Droplets are placed by rendering a disk, with additive blending enabled, into the liquid texture, which keeps the velocity \vec{v} and fluid volume q for all grid elements in its color channels. The velocity \vec{v} is updated for every time step Δt by calculating the new acceleration \vec{a} in tangent space by considering gravity and friction. The acceleration is derived from the sum of all external forces \vec{F} according to Newton's second law $\vec{F} = m \cdot \vec{a}$, and by considering the relationship $m = \rho \cdot q$ for homogeneous materials (with mass m , volume q , and density ρ): $\vec{a} = \frac{\vec{F}}{m} = \frac{\vec{F}}{\rho \cdot q}$.

In [JB09a] this method was later extended to handle dif-

ferent contact angles and to cope with non-planar surfaces such as facial meshes. Additionally, the authors presented an integration into the X3D standard for better usability. A physically-based method for simulating drops is presented in [WMT05], which achieves very realistic results and in contrast to other approaches also can simulate fluids dripping off a surface. This is realized by introducing the so-called virtual surface method for explicitly modeling fluid-solid interactions, but it requires several days for computation and is thus not suitable for interactive applications.

6. The Virtual Camera

Whereas skin color change and droplet flow to simulate effects like weeping were discussed in the previous chapter, in the rest of the paper rendering methods for lighting as well as a cinematographic camera approach to enhance these effects are described. Also, [dMP07, CO09] pointed out that not only framing but also the correct choice of lights, shadows, lenses, filters and similar effects are crucial for the final perception and can influence a user's perception of emotions. Hence, the environment can be used for communication, too, in that factors like lighting and camera control can be utilized to define, clarify, or emphasize a character's personality, role and interpersonal relations. Furthermore, as already was outlined in the introduction, simulating realistic lighting conditions is also essential for Mixed Reality scenarios, where the avatar and other virtual objects are inserted into a real scene.

Generating multimodal output such as natural language and graphics involves the coordination of all relevant modalities as well as the choice of vantage point from which to display the objects that are referred to linguistically [CO09]. Therefore, an MDS usually relies on default viewpoints from which suitable views of the elements of discourse can be achieved, and with a few exceptions up to now camera control has received little attention in computer graphics. Since capturing and communicating information by deciding where to position and how to move a camera in relation to the elements of the scene is a major concern of cinematography, Christie and Olivier [CO06] presented a state-of-the-art report where they discussed current approaches in automatic camera control while also analyzing main difficulties and challenges. Yet lenses, filters, and other visual effects, which are important for expressing moods or directing attention are left aside, though with recent GPUs they can be implemented rather efficiently in hardware.

6.1. Cinematographic Principles

Cinematography is the art of motion picture making, communicates the emotional state of a character, and includes aspects such as camera filters, lenses, framing, camera movement etc. A movie is made up of sequences of scenes that consist of several short shots (i.e. a continuous view that is filmed with one camera without interruption) [KM04]. Here,

it is important to maintain spatial and temporal continuity. Over the last century, filmmakers have developed a set of guidelines that allow involving the viewer in a story, and which nowadays are taken for granted [Haw04].

One such rule is the so-called "line of action" or "line of interest", which is either formed along the direction a single actor is facing or moving, or by the line connecting two interacting actors [Haw04]. The camera placement is specified relative to that imaginary line, which must not be crossed with any camera once it has been established in order to avoid confusion for the viewer. A common placement is for instance the so-called apex view, which frames two actors such that each of them is centered on one side of the screen [CAwH*96, HCS96], or an over-the-shoulder shot that already requires semantic knowledge. Cinematic terminology is derived from character oriented shot compositions, such as over-the-shoulder shots or close shots. These terms require a semantic rather than just geometric representation of objects. Furthermore, translating cinematographic notions into some sort of controller nodes is problematic, because even the seemingly simple notion of a shot encompasses many possible solutions [CO09]. However, high-level tools based on cinematic constructs can represent an advance over the existing keyframe-based methods.

Another important aspect concerns framing the scene, whereas generally the size of an object, mostly a human actor, is proportional to its present importance. Thus, some standard shot sizes like "full close-up", which shows the head and throat, or "full shot", which shows the full body by covering the whole screen vertically, have been established [CAwH*96, KM04]. Furthermore, there are also guidelines concerning the location of an object in the frame, which can be described by the so-called "rule of thirds". Changing the sizes of an object on the screen usually implies changing the camera distance. Therefore, occlusions can occur, which can be alleviated by modifying the camera's field of view instead. The target locations can be further modified by changing the camera angle relative to the line of action and floor plane [Haw04]. In this context, the problem of editing is distinguished from viewpoint planning [CO09].

Medium and close shots are classically used in dialogs in order to show the expressive parts of the body (gestures and mimics), whereas medium close-ups display the characters face, allowing to following the speech and to capture emotions. This way, even subtle effects like tears can be framed. Extreme close-ups can be used to emphasize the characters attitude or reaction, whilst focusing on the face and eyes [CO09]. So for example in a dialog, the height of the eyes is utilized to establish relationships between characters: eyes at the same level deliver a sense of balance between them and support the narrative in that way and vice versa. Because choosing specific types of camera placements etc. already requires semantical knowledge of the scene, e.g. in [CAwH*96, HCS96, JMAY03] it is distin-

guished between higher level concepts for modeling the cinematic expertise and lower level functionalities for designing a cinematographic camera module.

6.2. Cinematographic Approaches

In the context of film making, Blinn [Bli88] was one of the first to think about where to place a camera to get an interesting picture based on a description of the desired scene. In contrast to the traditional look-at transformation, he starts from the given 2D location (x_s, y_s) of the object of interest in screen space, and then solves for the appropriate camera matrix in 3D world space. A comprehensive introduction and overview on camera control in interactive applications including cinematographic considerations and high-level declarative approaches is given in [CO09]. And Hawkins [Haw04] explained how to incorporate filmmaking techniques, including the choice of lenses and filters to use, into games and similar applications.

Typically, in 3D applications viewpoints and camera animations are pre-defined such that important places are shown [Haw04]. Navigation then means to change the camera pose interactively. In [CAB*00] the Alice system is presented, which allows describing the interactive behavior of 3D objects more declaratively. In addition, in immersive VR environments special devices are needed, which are still uncommon, usually only available in one size, and in addition quite expensive [Rou00]. Moreover, users often have problems navigating through virtual environments, especially with typical 3D interaction devices. Therefore, [KM04] propose to apply the rules of filming also to games and other real-time 3D applications.

Recently, [KC08] presented an approach using hierarchical lines of action for generating correct camera setups for scenes that contain groups of more than two or three actors. Recently, [JB09c, WB09] proposed to adopt cinematographic techniques to X3D [Web08]. Whereas [JB09c] presented a set of nodes to ease framing dynamic content and including special effects, [WB09] discussed establishing higher level cinematographic concepts by introducing scene-graph nodes that encapsulate camera movements and shots.

Special visual effects (VFX) like depth-of-field [ST04, Ham07] and motion blur [Ros07] often directly result from the camera system used, and can be implemented on modern graphics hardware in a post-processing step. Depth-of-field sets the range of distances where all objects appear with acceptable sharpness. For implementation, [ST04] presented a GPU-based method that avoids blurry edges by also taking the depth buffer into account. After rendering the scene, the resulting image is filtered with a variable-sized filter kernel to simulate the circle of confusion, which is controlled by a depth-dependent blurriness factor that is based on the relative distance to the focus plane. As can be seen in Figure 10, depth-of-field for instance can be useful as an additional possibility to focus attention by blurring less important things.



Figure 10: *Depth-of-field (here shown with varying focal depth) is useful as an additional means to focus attention (compare [JWB*10]). ©2010 ACM.*

Furthermore, possible properties of the camera system need to be exposed to higher layers in a declarative way [JB09b, JWB*10]. Already [CAwH*96] stated, that interactive 3D applications fail to realize important storytelling capabilities and natural interactions with intelligent agents by ignoring the principles of cinematography. Hence, they formalize these principles into a declarative language for automatic camera control based on certain idioms (i.e. stereotypes for filming specific actions like a dialog scene). Likewise, [HCS96] organized such idioms with finite state machines on a higher level, whereas several camera modules with different behavior (e.g. 'follow' and 'apex') were responsible for the low-level camera placement.

A similar but knowledge-based approach that is implemented with the help of TVML for visualizing the scene via their TV production and simulation tool for desktop environments is proposed by [Hay98, JMAY03]. TVML is used to create animation scripts with predefined asset like characters etc., to create movements and dialogs, and to control special effects like lighting, camera and music, but it only allows predefined, non-interactive, linear stories and the synchronization of actions is limited. In [FGM02], intelligent camera control is utilized for visualizing important events in crowd simulations.

7. Set and Lighting

Lights, camera, lens filters, and shadows are essential factors for expressing moods, from a decoding perspective, and can influence the perception of emotions [dMP07]. For example, illuminating a face from below eye-level appears scary. Besides the virtual environment as such, the simulation of realistic lighting conditions is therefore also important.

7.1. Environment and Shadows

To consistently present a virtual character, in this section, we'll mainly discuss lighting aspects. Using the usually spherical image of the surrounding environment not only for lighting but also as the background image is a well-known method to improve the realistic look of a scene without an

increase in geometric complexity. To fit the lighting conditions of the environment map to those of the 3D scene, different techniques can be used, such as irradiance mapping [Kin05] or the extraction of light sources from real footage [GC03, ML07]. Automatic methods not only ease setting up more realistic lighting conditions as an important factor for expressing moods, but they are also rather useful in the context of Mixed Reality applications, where lighting reconstruction is still an open issue.

Accounting for complex lighting situations as given by the environment map generally comes along with high computational costs. A real-time approach to extract lights from HDR sphere maps for instance was also presented in [SS06] and [KSvA*06]. In [Deb05], a median-cut algorithm for extracting light sources from an environment map was proposed. This method later was adopted for real-time AR applications [ML07] in order to approximate the lighting situation via a finite set of light sources. Instead of employing image-based methods, which do not work well with the typically used subsurface scattering approximations for skin rendering (cp. Section 5.2.1), the aforementioned median-cut algorithm proposed by [Deb05, ML07] can be used for extracting the light sources, including their color and intensity, from the given environment maps.

This has the additional advantage that afterwards standard shadowing techniques can be applied. Real-time shadows are needed for depth cues and correct perception of a 3D scene [AMHH08, p. 331 ff.]. Whereas shadow volumes yield accurate shadow edges, they scale bad with size and complexity of a scene. Concerning shadow mapping techniques, we first have hard shadows like standard shadow mapping, which even runs on old graphics hardware but suffers from problems aliasing artifacts and imprecise depth comparisons. An in-depth survey of real-time hard shadow mapping methods, providing various improvements, recently was presented by [SWP10]. Parallel-split shadow maps especially focus on big scenes [ZSXL06]. The idea is to split-up the view frustum into several bins by using planes parallel to the viewing plane and then processing these parts individually.

Filtering approaches like the percentage closer filtering shadows (PCF) [RSC87] simulate soft shadows by calculating the mean value of n shadow tests. Likewise, variance shadow maps [DL06] tackle the problem of filtering depth textures by storing the mean and mean squared of the depth distribution to compute the variance over the filter region, which is then used to approximate the occlusion. Perspective PCF shadows (PCSS) allow for perspective correct soft shadows with a varying penumbra, which depends on the light source's size and its distance to an object [Fer05].

7.2. Lighting in Mixed Reality

Mixed Reality spans the whole continuum between VR and AR [JL04]. Whereas VR only deals with completely synthetic scenes, AR aims at integrating additional data into

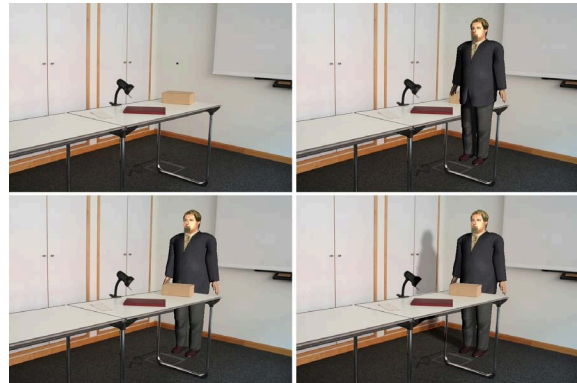


Figure 11: From top left to bottom right: real image of a meeting room; real scene augmented with virtual character (note that the character appears to be before the table); augmentation with additional occlusion handling (note that the character still seems to float on the floor); augmentation with occlusion and shadows.

real scenes. Especially in mobile computing in combination with geolocation-based services, there is a recent trend in augmenting the real world with virtual information, which is made possible due to increasing processing power, bandwidth, and 3D capabilities even on mobile devices. Thereby, fascinating new user experiences become possible, where e.g. a virtual character, as an augmented master teacher, explains and demonstrates the use of a newly bought device.

Hence, there also exist approaches for integrating virtual characters as human-computer interface, because such high-level context elements can more efficiently cope with augmented physical environments than e.g. explicit direct manipulation techniques via pointing devices or multi-touch. In this regard, Barakonyi and Schmalstieg [BS06] proposed a framework where virtual agents with autonomous behavior are employed as interface and interaction metaphor in the context of AR applications. One demonstration scenario here was a machine maintenance application, where a virtual animated repairman assisted an untrained user to maintain a real toy robot.

However, to augment the scene with plausible looking 3D objects, the question arises how to fit them as seamless as possible into real scenes. This at first requires that a live video of the real scene needs to be put behind virtual objects and thereby the exact pose of the user or camera somehow needs to be determined with vision-based tracking techniques. Though this is out of scope here (cf. e.g. [BWS06] for some more information on tracking), interfaces to integrate sensor data from external devices have to be considered, too [JFDB07]. Another problem is how to reconstruct and render the lighting fast enough. Therefore, the synthetic objects not only need to be registered geometrically but also photometrically for consistent lighting. A classification of illumination methods based on the amount of radiance and

geometry known from the real environment was presented by Jacobs and Loscos [JL04].

Lastly, the effects of changing the light transport paths of the real scene by inserting synthetic objects also have to be taken into account, like shadows cast from virtual onto real objects etc. In this regard, Figure 11 visualizes the basic problems here. Integrating virtual shadows into the real scene can be solved with differential rendering [Deb98], a rendering technique that adds the difference of two synthetic images to the real input image. Though this method was intended for offline rendering, [JFDB07] presented a real-time approach using shaders and multipass rendering. Generally, consistent illumination in AR is still an open field for research, because besides the lighting simulation, three other problems have to be solved in advance [GCHH04]: geometry reconstruction for handling e.g. occlusions; lighting reconstruction for recovering number and type of primary light sources; and material reconstruction for determining the reflectance properties of real materials, which is essential for computing correct inter-reflections and shadow color.

Although global illumination is too slow for real-time applications, it is still prevailing in the field of consistent illumination. Real-world lighting information for shading virtual objects was introduced by Debevec [Deb98]. He distinguishes between the local scene, for which a reflectance model is needed, and a distant scene, which only serves as the source for natural illumination. The incident scene radiance is captured using an omni-directional HDR image, the so-called light probe. In [SS06] image data is captured via a fish-eye lens or light probe. A GPU based real-time method was proposed in [GCHH04], in which the virtual object is illuminated by using an irradiance volume and cubic reflection maps. An approach combining geometric and photometric calibration was outlined in [PGL*06]. It uses a calibration object with known shape and normals and view independent albedo from which a light map is calculated.

Image based lighting is used to transfer real world lighting onto surfaces of virtual objects. Illumination of the real environment is passed to the virtual object via irradiance maps from a light probe, captured as HDR images. In [RH01], an image-based lighting technique with irradiance environment maps is presented. The authors use the spherical harmonic basis to filter spherical images and reuse the low-frequency spherical harmonic representation to simulate diffuse reflection via environment mapping. In [KDS04], ways of filtering environment maps are discussed to create different types of irradiance maps like diffuse and glossy ones.

In [SKS02], Sloan et al. introduce precomputed radiance transfer (PRT) with the spherical harmonics basis and describe the mathematical tools to precompute light transfer functions, simulating diffuse unshadowed, shadowed and self-reflected surfaces. Egges et al. [EPMT07, Pap06] presented a mixed reality framework for virtual characters that builds upon VHD++ [PPM*03]. For rendering, a dynamic

PRT (dPRT) approach is proposed, but it heavily relies on precomputation and manual work for defining special receiver and occluder objects, requires static lighting conditions, and also leaves subsurface scattering effects aside. A comparison of MR illumination models concerning their usability for deformable virtual humans is given in [Pap06].

8. Conclusions and Challenges

Interactive embodied conversational agents are an active field of research in e.g. Computer Graphics and especially Artificial Intelligence. The developments concerning visualization often concentrate on individual essential requirements, i.e. gestures, facial expression, and speech. Aspects like appearance and camera work are mostly ignored, although they help to correctly perceive the communicative intent. Moreover, for contextual relevance multimodal dialog systems need to be embedded into a broader environment, like in Mixed Reality applications, which requires a modular system design for all components. By then at the latest, rendering extrinsically gets important for achieving real mixed realities. In addition, certain effects that come along with various emotions, such as blushing and tears, intrinsically need to be displayed too. In this regard, expressive virtual characters also allow for novel evaluation possibilities. Hence, this report has reviewed relevant aspects of multimodal dialog systems, from high-level models to rendering issues, while it also defined a set of challenges, to support the proliferation of these systems.

As was shown with James Cameron's latest movie "Avatar", in this respect the so-called uncanny valley now can be overcome when sufficient money, time, and manpower are available. It therefore seems that comic-like figures as in the famous Pixar feature films can be avoided. However, even though the virtual characters in "Avatar" seem life-like, they are aliens with a blue skin, stylized faces, and different limbs living in a low-gravity world. Besides this obviously increasing interest in character technology, this film production also reveals the main challenges in this area for the coming years, since the demonstrated technologies here are far from real-time and the dialogs are not interactive at all. These issues are similar in the games industry, where every company has its own tools and engines. Though the results achieved are often very good, interactivity is restricted to scripted actions and the development of a new game title usually takes more than five years with average development costs of several million euros.

Acknowledgements

Part of this research has been carried out within the framework of the Excellence Cluster Multimodal Computing and Interaction (MMCI), sponsored by the German Research Foundation (DFG), and within the ANSWER project that has received funding from the European Community's Seventh Framework Programme.

Authors' CVs

Yvonne A. Jung received her PhD in early 2011 from Technical University Darmstadt, Germany. Between 2004 and 2010 she was a member of the Virtual and Augmented Reality group at the Fraunhofer Institute for Computer Graphics Research (IGD) in Darmstadt. After that, she changed to IGD's Visual Computing System Technologies group. She is also author of more than 30 scientific papers and serves as program committee member of the Web3D conference. Her research interests include GPU-based methods for real-time rendering and simulation, multimodal interaction techniques and virtual characters, as well as relighting in Mixed Reality.

Arjan Kuijper is lecturer at the TU Darmstadt and staff member of Fraunhofer IGD. He received a PhD in 2002 from the Department of Computer Science and Mathematics, Utrecht University, the Netherlands. January 2003 he became assistant research professor in The Image Group of the IT University of Copenhagen and November 2005 he started as senior researcher at the Johann Radon Institute for Computational and Applied Mathematics (RICAM), Linz, Austria. In 2009 he obtained his *venia legendi* at the Technical University of Graz, Austria, and since then he is also Privat Dozent at the Institute for Computer Graphics and Vision (ICG) at TU Graz and at the Interactive Graphics Systems Group at TU Darmstadt. He is author of over 80 peer-reviewed publications, and serves as reviewer for many journals and conferences, and as program committee member and organizer of conferences. His research interests cover all aspects of mathematics-based methods for computer vision and graphics.

Dieter W. Fellner is professor at the Interactive Graphics Systems Group at the TU Darmstadt and leads the Fraunhofer IGD in Darmstadt, Germany. He is also professor at the TU Graz, Austria, where he has established the Institute of Computer Graphics and Knowledge Visualization in 2006. The research activities there cover algorithms and software architectures to integrate modeling and rendering, efficient rendering and visualization algorithms, generative and reconstructive modeling, virtual and augmented reality, and digital libraries. His research projects comprise a broad spectrum of areas from formal languages, telematics services, and user interface design, to software engineering, computer graphics, and digital libraries. In the areas of computer graphics and digital libraries he is a member of the editorial boards of leading journals and a member of the program committees of many international conferences and workshops. In 1997 and 2003 he was program co-chair for the EUROGRAPHICS conference. Furthermore, he is an advisor for the German Scientific Council, the German Research Foundation, and the European Commission.

Michael Kipp is head of the EMBOTS (Embodied Agents) research group in the MMCI Cluster of Excellence, Saarbrücken. He is also senior researcher at the DFKI (German Research Center for AI). His group is conducting research and teaching in the area of intelligent embodied agents, in-

cluding human behavior analysis, automated character animation, intelligent interactive systems and formal evaluation. His research activities include initiating the first international workshop on sign language translation and avatar technology (SLTAT 2011), co-chairing the 9th International Conference on Intelligent Virtual Agents (IVA 2009) and being on the editorial board of the Journal on Multimodal User Interfaces (Springer). He is regularly serving as a program committee member on various international conferences including IVA, IUI, AAMAS and ACM Multimedia.

Jan Miksatko is a PhD candidate at Saarland University and member of the Embodied Agents group at DFKI and the MMCI Cluster of Excellence, Saarbrücken. He holds a MSc in Theoretical Computer Science, Charles University, Prague, and a MSc Computer Science, Kansas State University. His research interests are the development of multimodal user interfaces with embodied agents, evaluation of the innovative user interfaces and application of AI techniques into real systems in general.

Jonathan Gratch is an Associate Director for Virtual Humans Research at the University of Southern California's (USC) Institute for Creative Technologies, Research Associate Professor in the Department of Computer Science and co-director of USC's Computational Emotion Group. His research focuses on virtual humans and computational models of emotion. He studies the relationship between cognition and emotion, the cognitive processes underlying emotional responses, and the influence of emotion on decision making and physical behavior. He is on the editorial board of the journal *Emotion Review* and the President of the HUMAINE Association for Research on Emotions and Human-Machine Interaction. He is sitting member of the organizing committee for the International Conference on Intelligent Virtual Agents (IVA) and frequent organizer of conferences and workshops on emotion and virtual humans.

Daniel Thalmann is Professor at the Institute for Media Innovation at the Nanyang Technological University in Singapore. Until January 2011, he was Professor and Director of the Virtual Reality Lab (VRlab) at EPFL, Switzerland. He is a pioneer in research on Virtual Humans. His current research interests include real-time Virtual Humans in Virtual Reality, crowd simulation, and 3D interaction. He is coeditor-in-chief of the *Journal of Computer Animation and Virtual Worlds*, and member of the editorial board of 6 other journals. Daniel Thalmann was member of numerous Program Committees, Program Chair and CoChair of several conferences including IEEE VR, ACM VRST, and ACM VRCAI. He has also organized 5 courses at SIGGRAPH on human animation and crowd simulation. Additionally, he has published more than 500 papers in Graphics, Animation, and Virtual Reality. He is coeditor of 30 books, and coauthor of several books including "Crowd Simulation" and "Stepping into Virtual Reality". He received an Honorary Doctorate (Honoris Causa) from University Paul-Sabatier in Toulouse, France, in 2003, and the 2010 Eurographics Distinguished Career Award.

References

- [AB06] ARNAUD R., BARNES M.: *Collada*. AK Peters, 2006. 2
- [ABM00] ALEXA M., BEHR J., MÜLLER W.: The morph node. *Web3D - VRML 2000 Proceedings* (2000), 29–34. 8
- [ABTI*00] ANDRE E., BINSTED K., TANAKA-ISHII K., LUKE S., HERZOG G., RIST T.: Three robocup simulation league commentator systems. *AI Magazine* 21 (2000). 6
- [ADP06] AYLETT R. S., DIAS J., PAIVA A.: An affectively-driven planner for synthetic characters. vol. 1, pp. 2–10. 5, 6
- [AHJV08] AMOR H. B., HEUMER G., JUNG B., VITZTHUM A.: Grasp synthesis from low-dimensional probabilistic grasp models. *Journal of Visualization and Computer Animation* 19, 3-4 (2008), 445–454. 8
- [AMHH08] AKENINE-MÖLLER T., HAINES E., HOFFMANN N.: *Real-Time Rendering*, 3 ed. AK Peters, Wellesley, MA, 2008. 17
- [AR00] ANDRÉ E., RIST T.: Presenting Through Performing: On the Use of Multiple Lifelike Characters in Knowledge-Based Presentation Systems. In *Proceedings of the Second International Conference on Intelligent User Interfaces (IUI 2000)* (2000), pp. 1–8. 5
- [AR01] ANDRE E., RIST T.: Controlling the behavior of animated presentation agents in the interface: Scripting versus instructing. *AI Magazine* 22 (2001), 53–66. 5
- [ARL*10] ALEXANDER O., ROGERS M., LAMBETH W., CHIANG J.-Y., MA W.-C., WANG C.-C., DEBEVEC P.: The digital emily project: Achieving a photorealistic digital actor. *Computer Graphics and Applications, IEEE* 30, 4 (2010), 20–31. 8
- [BAC*06] BERTAILS F., AUDOLY B., CANI M.-P., QUERLEUX B., LEROY F., LÉVÊQUE J.-L.: Super-helices for predicting the dynamics of natural hair. *ACM Trans. Graph.* 25, 3 (2006), 1180–1187. Special issue: SIGGRAPH '06. 10
- [BAW10] BECKER-ASANO C., WACHSMUTH I.: Affective computing with primary and secondary emotions in a virtual human. *Autonom. Agents & Multi-Agent Systems* 20, 1 (2010), 32 – 49. 11, 13
- [BCN03] BANDO Y., CHEN B.-Y., NISHITA T.: Animating hair with loosely connected particles. *Computer Graphics Forum* 22, 3 (2003), 411 – 418. 10
- [BKMW05] BUNT H., KIPP M., MAYBURY M. T., WAHLSTER W.: Fusion and Coordination For Multimodal Interactive Information Presentation. In *Multimodal Intelligent Information Presentation*, Stock O., Zancanaro M., (Eds.). Kluwer, 2005, pp. 325–340. 4
- [BL03] BORSHUKOV G., LEWIS J. P.: Realistic human face rendering for "the matrix reloaded". In *SIGGRAPH '03: ACM SIGGRAPH 2003 Sketches & Applications* (New York, NY, USA, 2003), ACM, p. 1. 13
- [Bli88] BLINN J.: Where am I? What am I looking at? *IEEE Computer Graphics and Applications* 22 (1988), 179–188. 16
- [Bro91] BROOKS R. A.: Intelligence without representation. *Artificial Intelligence*, 47 (1991), 139–159. 6
- [BS06] BARAKONYI I., SCHMALSTIEG D.: Ubiquitous animated agents for augmented reality. In *ISMAR '06* (2006), pp. 145–154. 1, 17
- [BWB*95] BADLER N., WEBBER B., BECKET W., GEIB C., MOORE M., PELACHAUD C., REICH B., STONE M.: *Planning and parallel transition networks: Animation's new frontiers*. World Scientific, New York, 1995, pp. 101–117. 6
- [BWG10] BUISINE S., WANG Y., GRYSZPAN O.: Empirical investigation of the temporal relations between speech and facial expressions of emotion. *Journal on Multimodal User Interfaces* (2010), 1–8. 12
- [BWS06] BLESER G., WUEST H., STRICKER D.: Online camera pose estimation in partially known and dynamic scenes. In *ISMAR 2006 : Proc. of the Fourth IEEE and ACM Int. Symposium on Mixed and Augmented Reality* (Los Alamitos, Calif., 2006), IEEE Computer Society, pp. 56–65. 17
- [CAB*00] CONWAY M., AUDIA S., BURNETTE T., COSGROVE D., CHRISTIANSEN K.: Alice: lessons learned from building a 3d system for novices. In *CHI '00: Proceedings of SIGCHI conference on Human factors in computer systems* (NY, USA, 2000), ACM, pp. 486–493. 16
- [CAwH*96] CHRISTIANSON D. B., ANDERSON S. E., WEI HE L., SALESIN D., WELD D. S., COHEN M. F.: Declarative camera control for automatic cinematography. In *AAAI/IAAI, Vol. 1* (1996), pp. 148–155. 15, 16
- [CBC*00] CASSELL J., BICKMORE T., CAMPBELL L., VILHJÁLMSOHN H., YAN H.: *Human Conversation as a System Framework: Designing Embodied Conversational Agents*. MIT Press Cambridge, MA, USA, 2000, pp. 29–63. 6
- [CBCV00] CASSELL J., BICKMORE T., CAMPBELL L., VILHJÁLMSOHN H.: Human Conversation as a System Framework: Designing Embodied Conversational Agents. In *Embodied Conversational Agents*, Cassell J., Sullivan J., Prevost S., Churchill E., (Eds.). MIT Press, Cambridge, MA, 2000, pp. 29–63. 4
- [CBM09] COURGEON M., BUISINE S., MARTIN J.-C.: Impact of expressive wrinkles on perception of a virtual character's facial expressions of emotions. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents* (Berlin, Heidelberg, 2009), IVA '09, Springer-Verlag, pp. 201–214. 9
- [CCM02] CAVAZZA M., CHARLES F., MEAD S. J.: Interacting with virtual characters in interactive storytelling. vol. 1, ACM, pp. 318–325. 5
- [CCZB00] CHI D., COSTA M., ZHAO L., BADLER N.: The EMOTE Model for Effort and Shape. In *Proceedings of SIGGRAPH 2000* (2000), pp. 173–182. 8
- [CGV09] CAFARO A., GAITO R., VILHJÁLMSOHN H. H.: Animating idle gaze in public places. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents* (Berlin, Heidelberg, 2009), IVA '09, Springer-Verlag, pp. 250–256. 8
- [CLH*05] CHOSET H., LYNCH K. M., HUTCHINSON S., KANTOR G., BURGARD W., KAVRAKI L. E., THRUN S.: *Principles of Robot Motion: Theory, Algorithms, and Implementations*. MIT Press, Cambridge, MA, 2005. 9
- [CMJ08] COURGEON M., MARTIN J.-C., JACQUEMIN C.: Marc: a multimodal affective and reactive character. In *Proceedings of the Workshop on Affective Interaction on Natural Environment (AFFINE)* (2008). 12
- [CO06] CHRISTIE M., OLIVIER P.: Camera Control in Computer Graphics. In *EG 2006 - State of the Art Reports* (2006), Eurographics Association, pp. 89–113. 3, 15
- [CO09] CHRISTIE M., OLIVIER P.: Camera control in computer graphics: models, techniques and applications. In *SIGGRAPH ASIA '09: ACM SIGGRAPH ASIA 2009 Courses* (New York, NY, USA, 2009), ACM, pp. 1–197. 2, 15, 16
- [CSPC00] CASSELL J., SULLIVAN J., PREVOST S., CHURCHILL E.: *Embodied Conversational Agents*. MIT Press, Cambridge, MA, 2000. 4
- [CT00] CAICEDO A., THALMANN D.: Virtual humanoids: Let them be autonomous without losing control. 6

- [CVB01] CASSELL J., VILHJÁLMSSON H. H., BICKMORE T.: Beat: the behavior expression animation toolkit. *ACM*, pp. 477–486. 7
- [DCPPdR01] DE CAROLIS B., PELACHAUD C., POGGI I., DE ROSIS F.: Behavior planning for a reflexive agent. In *Proceedings of the 17th international joint conference on Artificial intelligence - Volume 2* (San Francisco, CA, USA, 2001), Morgan Kaufmann Publishers Inc., pp. 1059–1064. 6
- [DCPPS02] DE CAROLIS B., PELACHAUD C., POGGI I., STEEDMAN M.: Apm1, a mark-up language for believable behavior generation. In *Embodied conversational agents - let's specify and evaluate them! Proc. of AAMAS Workshop* (2002). 4, 7
- [dCPS04] DE CAROLIS B., PELACHAUD C., POGGI I., STEEDMAN M.: *APML, a mark-up language for believable behavior generation*. Springer, 2004, pp. 65–85. 7
- [DdJP09] DIJK C., DE JONG P., PETERS M.: The remedial value of blushing in the context of transgressions and mishaps. *Emotion* 9, 2 (2009), 287–291. 12
- [Deb98] DEBEVEC P.: Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proc. of SIGGRAPH 98* (1998), CG Proc., Annual Conf. Series, pp. 189–198. 18
- [Deb05] DEBEVEC P.: A median cut algorithm for light probe sampling. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Posters* (New York, USA, 2005), ACM, p. 66. 17
- [DHT*00] DEBEVEC P., HAWKINS T., TCHOU C., DUIKER H.-P., SAROKIN W., SAGAR M.: Acquiring the reflectance field of a human face. In *SIGGRAPH '00* (New York, NY, USA, 2000), ACM Press/Addison-Wesley Publishing Co., pp. 145–156. 8
- [DL06] DONNELLY W., LAURITZEN A.: Variance shadow maps. In *13D '06: Proc. of 2006 symposium on Interactive 3D graphics and games* (New York, USA, 2006), ACM, pp. 161–165. 17
- [dL07] D'EON E., LUEBKE D.: Advanced techniques for realistic real-time skin rendering. In *GPU Gems 3*. Addison-Wesley, 2007, ch. 14, pp. 293–347. 13, 14
- [dMG09] DE MELO C. M., GRATCH J.: Expression of emotions using wrinkles, blushing, sweating and tears. In *Intelligent Virtual Agents: 9th International Conference, IVA 2009* (Heidelberg, 2009), Springer, pp. 188–200. 12
- [dMP07] DE MELO C., PAIVA A.: Expression of emotions in virtual humans using lights, shadows, composition and filters. In *ACII '07* (2007), pp. 546–557. 15, 16
- [DMTKT93] DALDEGAN A., MAGNENAT-THALMANN N., KURIHARA T., THALMANN D.: An integrated system for modeling, animating and rendering hair. *Computer Graphics Forum* 12, 3 (1993), 211–221. 10
- [DS03] DACHSBACHER C., STAMMINGER M.: Translucent shadow maps. pp. 197–201. *Proceedings of Eurographics Symposium on Rendering*. 13
- [dST05] DE SEVIN E., THALMANN D.: A motivational model of action selection for virtual humans. *Computer Graphics International* (2005), 213–220. 6
- [DvM00] DEHN D. M., VAN MULKEN S.: The impact of animated interface agents: a review of empirical research. *Int. J. Hum.-Comput. Stud.* 52, 1 (2000), 1–22. 4
- [EB05] ELLIS P. M., BRYSON J. J.: The significance of textures for affective interfaces. 394–404. 12
- [Eds03] EDSALL J.: Animation blending: Achieving inverse kinematics and more. *Gamasutra*, 2003. http://www.gamasutra.com/features/20030704/edsall_01.shtml. 9
- [EF78] EKMAN P., FRIESEN W. V.: The facial action coding system. *Consulting Psychologists' Press* (1978). 11
- [Egg06] EGGES A.: *Real-time Animation of Interactive Virtual Humans*. Dissertation, University of Geneva, 2006. 9
- [EPMT07] EGGES A., PAPAGIANNAKIS G., MAGNENAT-THALMANN N.: Presence and interaction in mixed reality environments. *Vis. Comput.* 23, 5 (2007), 317–333. 18
- [Fer05] FERNANDO R.: Percentage-closer soft shadows. In *SIGGRAPH '05: Sketches* (New York, USA, 2005), ACM Press, p. 35. 17
- [Fer10] FERRÉ G.: Timing relationships between speech and co-verbal gestures in spontaneous french. In *Proc. of the workshop on "Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality"* (2010), pp. 86–90. 4
- [FGM02] FERREIRA F. P., GELATTI G., MUSSE S. R.: Intelligent virtual environment and camera control in behavioural simulation. In *SIBGRAPI '02: Proc. of the 15th Brazilian Symp. on CG and Image Proc.* (Washington, DC, USA, 2002), IEEE, pp. 365–372. 16
- [GAP*10] GOBRON S., AHN J., PALTOGLOU G., THELWALL M., THALMANN D.: From sentence to emotion: a real-time three-dimensional graphics metaphor of emotions extracted from text. *Vis. Comput.* 26, 6-8 (2010), 505–519. 13
- [GC03] GIBSON S., CHALMERS A.: Photorealistic augmented reality. *Eurographics 2003 Tutorial*, 2003. 17
- [GCHH04] GIBSON S., COOK J., HOWARD T., HUBBOLD R.: *ARIS: Augmented Reality Image Synthesis*. Tech. Rep. IST-2000-28707, University Manchester, 2004. 3, 18
- [GCM97] GRAND S., CLIFF D., MALHOTRA A.: Creatures: artificial life autonomous software agents for home entertainment. *ACM*, pp. 22–29. 6
- [Geb05] GEBHARD P.: ALMA - a layered model of affect. In *Proc. of the Fourth Int. Joint Conference on Autonomous Agents and Multiagent Systems* (2005), pp. 29–36. 13
- [GKKR03] GEBHARD P., KIPP M., KLESEN M., RIST T.: Authoring scenes for adaptive, interactive performances. *Proc. of the 2nd Int'l Joint Conf. on Autonomous agents and multiagent systems* (2003), 725–732. 6
- [Gle98] GLEICHER M.: Retargeting motion to new characters. In *SIGGRAPH '98* (New York, USA, 1998), ACM, pp. 33–42. 8
- [GM04] GRATCH J., MARSELLA S.: Evaluating the modeling and use of emotion in virtual humans. pp. 320–327. 6
- [GMMT*07] GIACOMO T. D., MOCCOZET L., MAGNENAT-THALMANN N., BOULIC R., THALMANN D.: Towards Automatic Character Skeletonization and Interactive Skin Deformation. In *EG 2007 - State of the Art Reports* (2007), pp. 47–61. 8
- [GRA*02] GRATCH J., RICKEL J., ANDRE E., CASSELL J., PETAJAN E., BADLER N.: Creating interactive virtual humans: some assembly required. *Intelligent Systems, IEEE* 17, 4 (2002), 54–63. 2, 7
- [Gre04] GREEN S.: Real-time approximations to subsurface scattering. In *GPU Gems*. Addison Wesley, 2004, ch. 16, pp. 263–278. 13, 14
- [GS10] GILLIES M., SPANLANG B.: Comparing and evaluating real time character engines for virtual environments. *Presence: Teleoper. Virtual Environ.* 19, 2 (2010), 95–117. 7, 8, 9
- [GSI*04] GÖBEL S., SCHNEIDER O., IURGEL I., FEIX A., KNÖPFLE C., RETTIG A.: Virtual human: Storytelling and computer graphics for a virtual human platform. In *TIDSE 2004* (Darmstadt, 2004), pp. 79–88. 9

- [Gul04] GULZ A.: Benefits of virtual characters in computer based learning environments: Claims and evidence. *Int. J. Artif. Intell. Ed.* 14, 3,4 (2004), 313–334. 4
- [GVT00] GUYE-VUILLÈME A., THALMANN D.: A high-level architecture for believable social agents. *Virtual Reality* 5 (2000), 95–106. 6
- [GVT08] GUTIERREZ M. A., VEXO F., THALMANN D.: *Stepping into Virtual Reality*. Springer, London, 2008. 1
- [Ham06] HAMM A.: Psychologie der Emotionen. In *Neuropsychologie*, Karnath H., Thier P., (Eds.). Springer Verlag, Berlin, Heidelberg, 2006, pp. 527–535. 11
- [Ham07] HAMMON E.: Practical post-process depth of field. In *GPU Gems 3*, Nguyen H., (Ed.). Addison-Wesley, 2007, ch. 28, pp. 583–605. 16
- [Har87] HAREL D.: Statecharts: A visual formalism for complex systems. *Science of Computer Programming* 8 (1987), 231–274. 6
- [Haw04] HAWKINS B.: *Real-Time Cinematography for Games*. Charles River Media Inc., Rockland, USA, 2004. 15, 16
- [Hay98] HAYASHI M.: TVML (TV program making language). In *SIGGRAPH '98: ACM SIGGRAPH 98 Conference abstracts and applications* (New York, USA, 1998), ACM, p. 292. 16
- [HCS96] HE L.-W., COHEN M. F., SALESIN D. H.: The virtual cinematographer: a paradigm for automatic real-time camera control and directing. In *SIGGRAPH '96: Proceedings* (New York, USA, 1996), ACM, pp. 217–224. 15, 16
- [Her03] HERY C.: Implementing a skin bssrdf, 2003. RenderMan course notes, Siggraph. 13
- [HEV03] HUANG Z., ELIÈNS A., VISSER C.: Implementation of a scripting language for vml/x3d-based embodied agents. In *Web3D '03: Proc. of the 8th int. conf. on 3D Web technology* (NY, USA, 2003), ACM, pp. 91–100. 7
- [HG41] HENYEL L., GREENSTEIN J. L.: Diffuse radiation in the galaxy. *Astrophysics J.* 93 (1941), 70–83. 13
- [HG07] HELOIR A., GIBET S.: A qualitative and quantitative characterization of style in sign language gestures. In *Gesture in Human-Computer Interaction and Simulation, 7th International Gesture Workshop, GW 2007* (2007). 8
- [HJGP08] HAJJAR J.-F. E., JOLIVET V., GHAZANFARPOUR D., PUEYO X.: A model for real-time on-surface flows. *The Visual Computer* 25, 2 (2008), 87–100. 14
- [HK93] HANRAHAN P., KRÜGER W.: Reflection from layered surfaces due to subsurface scattering. pp. 165–174. Proceedings in SIGGRAPH 93. 13
- [HK09] HELOIR A., KIPP M.: EMBR – a realtime animation engine for interactive embodied agents. In *Intelligent Virtual Agents, IVA 2009* (Heidelberg, 2009), Springer, pp. 393–404. 2, 5, 9
- [HK10] HELOIR A., KIPP M.: Realtime animation of interactive agents: Specification and realization. *Journal of Applied Artificial Intelligence* 24, 6 (2010), 510–529. 7, 9
- [HKM*08] HEYLEN D., KOPP S., MARSELLA S. C., PELACHAUD C., VILHJALMSSON H.: The next step towards a function markup language. Springer. 7
- [HMP06] HARTMANN B., MANCINI M., PELACHAUD C.: Implementing Expressive Gesture Synthesis for Embodied Conversational Agents. In *Gesture in Human-Computer Interaction and Simulation* (2006). 8
- [HMT01] HADAP S., MAGNENAT-THALMANN N.: Modeling dynamic hair as a continuum. *Computer Graphics Forum* 20, 3 (2001). 10
- [HR06] HERZOG G., REITHINGER N.: The smartkom architecture: A framework for multimodal dialogue systems. In *SmartKom: Foundations of Multimodal Dialogue Systems*, Wahlster W., (Ed.), Cognitive Technologies. Springer, Berlin, 2006, pp. 55–70. 2, 5
- [HWBF95] HODGINS J. K., WOOTEEN W. L., BROGAN D. C., F.O'BRIEN J.: Animating human athletics. In *Proceedings of ACM SIGGRAPH'95* (Aug. 1995), Computer Graphics Proceedings, Annual Conference Series, pp. 71–78. 8
- [iAUK92] ICHI ANJYO K., USAMI Y., KURIHARA T.: A simple method for extracting the natural beauty of hair. In *SIGGRAPH '92* (New York, USA, 1992), ACM, pp. 111–120. 10
- [IC05] IERONUTTI L., CHITTARO L.: A virtual human architecture that integrates kinematic, physical and behavioral aspects to control h-anim characters. In *Web3D '05: Proc. of the 10th int. conf. on 3D Web technology* (NY, USA, 2005), ACM, pp. 75–83. 7
- [INN05] IGARASHI T., NISHINO K., NAYAR S.: *The Appearance of Human Skin*. Tech. rep., Department of Computer Science, Columbia University, 2005. 13
- [IR10] IR: Instant Reality, 2010. <http://doc.instantreality.org/documentation/nodetype/>. 9
- [JB08] JUNG Y., BEHR J.: Extending H-Anim and X3D for advanced animation control. In *Proceedings Web3D 2008* (New York, USA, 2008), Spencer S., (Ed.), ACM, pp. 57–65. 2, 7, 9
- [JB09a] JUNG Y., BEHR J.: GPU-based real-time on-surface droplet flow in X3D. In *Web3D 2009* (New York, USA, 2009), Spencer S., (Ed.), ACM, pp. 51–54. 14
- [JB09b] JUNG Y., BEHR J.: Simplifying the integration of virtual humans into dialog-like VR systems. In *Proc. of IEEE Virtual Reality 2009 Workshop: 2nd SEARIS Workshop* (2009), Shaker, pp. 41–50. 7, 9, 16
- [JB09c] JUNG Y., BEHR J.: Towards a new camera model for x3d. In *Proceedings Web3D 2009* (New York, USA, 2009), Spencer S., (Ed.), ACM Press, pp. 79–82. 16
- [JFDB07] JUNG Y., FRANKE T., DÄHNE P., BEHR J.: Enhancing X3D for advanced MR appliances. In *Proceedings Web3D '07* (NY, USA, 2007), ACM Press, pp. 27–36. 17, 18
- [JH02] JONSSON M., HAST A.: Animation of water droplet flow on structured surfaces. In *SIGRAD2002* (Linköping, Sweden, 2002), Linköping University Press, pp. 17–22. 14
- [JK03] JUNG B., KOPP S.: Flurmax: An interactive virtual agent for entertaining visitors in a hallway. In *IVA* (2003), pp. 23–26. 5, 7, 9
- [JK05] JUNG Y., KNÖPFLE C.: Styling and real-time simulation of human hair. In *Intelligent Technologies for Interactive Entertainment. Proceedings: First International Conference, INTERTAIN 2005* (Heidelberg, 2005), Maybury M., Stock O., Wahlster W., (Eds.), Springer, pp. 240–245. 10
- [JK06] JUNG Y., KNÖPFLE C.: Dynamic aspects of real-time face-rendering. In *Proceedings VRST Cyprus 2006* (New York, 2006), ACM, pp. 193–196. 3, 12, 13, 14
- [JK07] JUNG Y., KNÖPFLE C.: Real-time rendering and animation of virtual characters. *The International Journal of Virtual Reality (IJVR)* 6, 4 (2007), 55–66. 9
- [JL04] JACOBS K., LOSCOS C.: Classification of Illumination Methods for Mixed Reality. In *EG 2004 - State of the Art Reports* (2004), Eurographics Association, pp. 95–118. 17, 18
- [JMAY03] JINHONG S., MIYAZAKI S., AOKI T., YASUDA H.: Filmmaking production system with rule-based reasoning. In *Image and Vision Computing* (New Zealand, 2003), pp. 366–371. 15, 16

- [JMLH01] JENSEN H. W., MARSCHNER S. R., LEVOY M., HANRAHAN P.: A practical model for subsurface light transport. In *SIGGRAPH '01* (New York, USA, 2001), ACM, pp. 511–518. [13](#)
- [JRKL05] JUNG Y., RETTIG A., KLAR O., LEHR T.: Realistic real-time hair simulation and rendering. In *Vision, Video, and Graphics. Proceedings 2005* (Aire-la-Ville, 2005), Trucco E., Chantler M., (Eds.), Eurographics Association, pp. 229–236. [10](#)
- [JRS*07] JOHNSEN K., RAJ A., STEVENS A., LIND D. S., LOK B.: The validity of a virtual human experience for interpersonal skills education. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, USA, 2007), ACM, pp. 1049–1058. [1](#)
- [JW10] JUNG Y., WAGNER S.: Emotional factors in face rendering. In *IADIS Multi Conference on Computer Science and Information Systems 2010: Proceedings IADIS Interfaces and Human Computer Interaction* (2010), IADIS Press. 354–358. [13](#)
- [JWB*10] JUNG Y., WAGNER S., BEHR J., JUNG C., FELLNER D. W.: Storyboarding and pre-visualization with x3d. In *Proceedings Web3D 2010: 15th International Conference on 3D Web Technology* (New York, USA, 2010), Spencer S., (Ed.), ACM Press, pp. 73–81. [16](#)
- [JWKF09] JUNG Y., WEBER C., KEIL J., FRANKE T.: Real-time rendering of skin changes caused by emotions. In *Intelligent Virtual Agents: 9th International Conference, IVA 2009* (Heidelberg, 2009), Springer, pp. 504–505. [12](#)
- [JWSG10] JIMENEZ J., WHELAN D., SUNDSTEDT V., GUTIERREZ D.: Real-time realistic skin translucency. *Computer Graphics and Applications, IEEE* 30, 4 (2010), 32–41. [14](#)
- [KB04] KRISHNASWAMY A., BARANOSKI G. V. G.: A biophysically-based spectral model of light interaction with human skin. *Computer Graphics Forum* 23, 3 (2004), 331–340. [13](#)
- [KC08] KARDAN K., CASANOVA H.: Virtual cinematography of group scenes using hierarchical lines of actions. In *Sandbox '08: Proc. of symp. on Video games* (NY, USA, 2008), ACM, pp. 171–178. [16](#)
- [KDD*05] KUPPEVELT J. V., DYBKJ&AELIG;R L., DYBKJ&AELIG;R L., BERNSEN N. O., WILKS Y., WEBB N., SETZER A., CATIZONE R.: *Advances in Natural Multimodal Dialogue Systems*. 2005. [4](#)
- [KDM10] KUBO H., DOBASHI Y., MORISHIMA S.: Curvature-dependent reflectance function for rendering translucent materials. In *SIGGRAPH '10: ACM SIGGRAPH 2010 Talks* (New York, NY, USA, 2010), ACM, pp. 1–1. [14](#)
- [KDS04] KAUTZ J., DAUBERT K., SEIDEL H.-P.: Advanced environment mapping in vr applications. *Computers & Graphics* 28, 1 (2004), 99–104. [18](#)
- [Ken04] KENDON A.: *Gesture – Visible Action as Utterance*. Cambridge University Press, Cambridge, 2004. [4](#)
- [KG07] KLESEN M., GEBHARD P.: Affective multimodal control of virtual characters. *IJVR* 6, 4 (2007), 43–54. [9](#), [12](#), [13](#)
- [KG08] KIPP M., GEBHARD P.: Igaze: Studying reactive gaze behavior in semi-immersive human-avatar interactions. In *Proceedings of the 8th International Conference on Intelligent Virtual Agents (IVA-08)* (2008). [8](#)
- [KGP02] KOVAR L., GLEICHER M., PIGHIN F.: Motion graphs. *ACM Trans. Graph.* 21, 3 (2002), 473–482. [9](#)
- [KHGS10] KIPP M., HELOIR A., GEBHARD P., SCHRÖDER M.: Realizing multimodal behavior: Closing the gap between behavior planning and embodied agent presentation. In *Intelligent Virtual Agents: 10th Intl. Conf., IVA 2010* (2010), Springer. [2](#), [3](#), [5](#), [7](#)
- [KHS04] KOSTER M., HABER J., SEIDEL H.-P.: Real-time rendering of human hair using programmable graphics hardware. In *CGI '04: Proceedings of the Computer Graphics International* (2004), IEEE Computer Society, pp. 248–256. [10](#)
- [Kin05] KING G.: Real-time computation of dynamic irradiance environment maps. In *GPU Gems 2*, Pharr M., (Ed.). Addison Wesley, 2005, ch. 10, pp. 167–176. [17](#)
- [KIY99] KANEDA K., IKEDA S., YAMASHITA H.: Animation of water droplets moving down a surface. *The Journal of Visualization and Computer Animation* 10, 1 (1999), 15–26. [14](#)
- [KK89] KAJIYA J. T., KAY T. L.: Rendering fur with three dimensional textures. In *SIGGRAPH '89* (1989), ACM Press, pp. 271–280. [10](#)
- [KKGR03] KLESEN M., KIPP M., GEBHARD P., RIST T.: Staging exhibitions: methods and tools for modeling narrative structure to produce interactive performances with virtual actors. *Virtual Reality. Special Issue on Storytelling in Virtual Environments* 7 (2003), 17–29. [6](#)
- [KKM*06a] KOPP S., KRENN B., MARSELLA S., MARSHALL A., PELACHAUD C., PIRKER H., THÓRISSON K., VILHJÁLMS-SON H.: Towards a common framework for multimodal generation: The behavior markup language. In *IVA 06* (2006), pp. 205–217. [4](#), [5](#)
- [KKM*06b] KOPP S., KRENN B., MARSELLA S., MARSHALL A. N., PELACHAUD C., PIRKER H., THÓRISSON K. R., VILHJÁLMS-SON H.: Towards a common framework for multimodal generation: The behavior markup language. In *Proc. of IVA-06* (2006). [5](#), [7](#)
- [KKW02] KRANSTEDT A., KOPP S., WACHSMUTH I.: Murml: A multimodal utterance representation markup language for conversational agents. In *Embodied conversational agents - let's specify and evaluate them! Proc. of AAMAS Workshop* (2002). [7](#), [9](#)
- [KKY93] KANEDA K., KAGAWA T., YAMASHITA H.: Animation of water droplets on a glass plate. *Proceedings of Computer Animation* 93 (1993), 177–189. [14](#)
- [KM04] KNEAFSEY J., MCCABE H.: Camera control through cinematography for virtual environments: A state of the art report. In *Proc. of Eurographics Ireland Chapter Workshop 2004* (2004). [15](#), [16](#)
- [KMT94] KALRA P., MAGNENAT-THALMANN N.: Modeling of vascular expressions. In *Computer Animation '94* (Geneva, 1994), pp. 50–58. [11](#), [12](#)
- [KN01] KIM T.-Y., NEUMANN U.: Opacity shadow maps. In *Proc. of the 12th Eurographics Workshop on Rendering Techniques* (London, UK, 2001), Springer-Verlag, pp. 177–182. [10](#)
- [KN02] KIM T.-Y., NEUMANN U.: Interactive multiresolution hair modeling and editing. In *SIGGRAPH '02* (2002), ACM Press, pp. 620–629. [10](#), [11](#)
- [KNKA07] KIPP M., NEFF M., KIPP K. H., ALBRECHT I.: Toward Natural Gesture Synthesis: Evaluating gesture units in a data-driven approach. In *Proc. of the 7th International Conference on Intelligent Virtual Agents (IVA-07)* (2007), Springer, pp. 15–28. [8](#)
- [KP03] KOENDERINK J., PONT S.: The secret of velvety skin. *Machine Vision and Applications* 14, 4 (2003), 260–268. [13](#)
- [KP04] KRENN B., PIRKER H.: Defining the gesticon: Language and gesture coordination for interacting embodied agents. In *Proc. AISB'04 Symp. on Language, Speech and Gesture for Expressive Characters* (Leeds, UK, 2004), pp. 107–115. [7](#)

- [KSVA*06] KORN M., STANGE M., VON ARB A., BLUM L., KREIL M., KUNZE K.-J., ANHENN J., WALLRATH T., GROSCH T.: Interactive augmentation of live images using a hdr stereo camera. 17
- [LC87] LORENSEN W. E., CLINE H. E.: Marching cubes: A high resolution 3d surface construction algorithm. In *SIGGRAPH '87* (1987), pp. 163–169. 14
- [LCK*97] LESTER J. C., CONVERSE S. A., KAHLER S. E., BARLOW S. T., STONE B. A., BHOGAL R.: The persona effect: Affective impact of animated pedagogical agents. In *Proceedings of CHI'97 Human Factors in Computing Systems* (New York, 1997), ACM Press, pp. 359–366. 4
- [LCR*02] LEE J., CHAI J., REITSMA P. S. A., HODGINS J. K., POLLARD N. S.: Interactive control of avatars animated with human motion data. *ACM Trans. Graph.* 21, 3 (2002), 491–500. 9
- [LKTK10] LEVINE S., KRÄHENBÜHL P., THRUN S., KOLTUN V.: Gesture controllers. In *Proc. of SIGGRAPH* (2010). 8
- [LM06] LEE J., MARSELLA S.: Nonverbal behavior generator for embodied conversational agents. In *Proc. of the 6th International Conference on Intelligent Virtual Agents* (2006), Springer, pp. 243–255. 8
- [LPR07] LÖCKELT M., PFLEGER N., REITHINGER N.: Multi-party conversation for mixed reality. *IJVR* 6, 4 (2007), 31–42. 2, 5
- [LTK09] LEVINE S., THEOBALT C., KOLTUN V.: Real-time prosody-driven synthesis of body language. In *Proc. of SIGGRAPHAsia* (2009). 8
- [LVTC97] LESTER J. C., VOERMAN J. L., TOWNS S. G., CALLAWAY C. B.: Cosmo: A life-like animated pedagogical agent. *Proceedings of IJCAI97 Workshop on Animated Interface Agents* (1997). 5
- [Mar96] MARIAZOUOLS C.: *Psychophysiologie von Scham und Erröten*. Dissertation, Ludwig-Maximilians-Universität München, 1996. 11
- [Mar01] MARRIOTT A.: VHML, 2001. <http://www.vhml.org/>. 7
- [McN05] MCNEILL D.: *Gesture and Thought*. University of Chicago Press, Chicago, 2005. 4
- [McT99] MCTEAR M. F.: Software to support research and development of spoken dialogue systems. In *Proceedings of Eurospeech* (1999), pp. 339–342. 6
- [MGHK09] MACDORMAN K. F., GREEN R. D., HO C.-C., KOCH C. T.: Too real for comfort? uncanny responses to computer generated faces. *Computers in Human Behavior* 25, 3 (2009), 695–710. 12
- [MGP] MARSELLA S., GRATCH J., PETTA P.: Computational models of emotion. In *A blueprint for an affectively competent agent: Cross-fertilization between Emotion Psychology, Affective Neuroscience, and Affective Computing*, Scherer K., Bänziger T., Roesch E., (Eds.), Oxford University Press. In press. 12
- [MJC*03] MARSCHNER S. R., JENSEN H. W., CAMMARANO M., WORLEY S., HANRAHAN P.: Light scattering from human hair fibers. *ACM Trans. Graph.* 22, 3 (2003), 780–791. 10
- [MK09] MIKSATKO J., KIPP M.: Hybrid control for embodied agents applications. In *Proc. of the 32nd Annual Conference on Artificial Intelligence* (2009), Mertsching B., (Ed.). 6
- [MKB*03] MERTENS T., KAUTZ J., BEKAERT P., REETH F. V., SEIDEL H.-P.: Efficient rendering of local subsurface scattering. In *PG '03: Proc. of the 11th Pacific Conf. on CG and Applications* (Washington, DC, USA, 2003), IEEE Computer Society, pp. 51–58. 13
- [MKK10] MIKSATKO J., KIPP K. H., KIPP M.: The persona zero-effect: Evaluating virtual character benefits on a learning task. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents (IVA-10)*. Springer, 2010. 4
- [ML07] MADSEN C. B., LAURSEN R. E.: A scalable gpu-based approach to shading and shadowing for photorealistic real-time augmented reality. In *GRAPP 2007* (2007), pp. 252–261. 17
- [Mor70] MORI M.: The uncanny valley. *Energy* 7, 4 (1970), 33–35. 4
- [MP07] MCCANN J., POLLARD N.: Responsive characters from motion fragments. *ACM Trans. Graph.* 26, 3 (2007), 6. 9
- [MS03] MATEAS M., STERN A.: Facade: An experiment in building a fully-realized interactive drama, 2003. 5
- [MTE06] MAGNENAT-THALMANN N., EGGES A.: Interactive virtual humans in real-time virtual environments. *IJVR* 5, 2 (2006), 15–24. 9
- [MV02] MOUNDRIDOU M., VIRVOU M.: Evaluating the persona effect of an interface agent in a tutoring system. *Journal of Computer Assisted Learning* 18 (Sept. 2002), 253–261. 4
- [Nar00] NAREYEK A.: Intelligent agents for computer games. *Computers and Games, Second International Conference, CG 2063* (2000), 414?422. 5
- [NB97] NOMA T., BADLER N.: A virtual human presenter. *Proc. of the IJCAI Workshop on Animated Interface Agents: Making Them Intelligent 4551* (1997). 5
- [NBA09] N. BEE S. F., ANDRÉ E.: Relations between facial display, eye gaze and head tilt: Dominance perception variations of virtual agents. In *Proceedings of Affective Computing and Intelligent Interaction (ACII)* (2009). 8
- [NBMP09] NIEWIADOMSKI R., BEVACQUA E., MANCINI M., PELACHAUD C.: Greta: an interactive expressive eca system. In *AAMAS* (2) (2009), pp. 1399–1400. 2
- [ND05] NGUYEN H., DONNELLY W.: Hair animation and rendering in the nalu demo. In *GPU Gems 2*. Addison Wesley, 2005, ch. 23, pp. 361–380. 9, 10, 11
- [NF02] NEFF M., FIUME E.: Modeling tension and relaxation for computer animation. In *SCA '02: Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation* (New York, NY, USA, 2002), ACM, pp. 81–88. 8
- [NF05] NEFF M., FIUME E.: Aer: Aesthetic exploration and refinement for expressive character animation. In *ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2005). 8
- [NKAS08] NEFF M., KIPP M., ALBRECHT I., SEIDEL H.-P.: Gesture Modeling and Animation Based on a Probabilistic Recreation of Speaker Style. *ACM Transactions on Graphics* 27, 1 (March 2008), 1–24. 8
- [NTSP02] N.MAGNENAT-THALMANN, S.HADAP, P.KALRA: State of the art in hair simulation. In *Int. Workshop on Human Modeling and Animation, Korea Computer Graphics Society* (2002), pp. 3–9. 10
- [NWA10] NEFF M., WANG Y., ABBOTT R., WALKER M.: Evaluating the effect of gesture and language on personality perception in conversational agents. In *Proceedings of the 10th international conference on Intelligent virtual agents* (Berlin, Heidelberg, 2010), IVA'10, Springer-Verlag, pp. 222–235. 8
- [OCC88] ORTONY A., CLORE G., COLLINS A.: *The Cognitive Structure of Emotions*. Cambridge University Press, 1988. 11, 13
- [OODPC*09] OYARZUN D., ORTIZ A., DEL PUY CARRETERO

- M., GELISSEN J., GARCIA-ALONSO A., SIVAN Y.: Adml: a framework for representing inhabitants in 3d virtual worlds. In *Web3D '09* (New York, NY, USA, 2009), ACM, pp. 83–90. 7
- [OPV*99] O'TOOLE A. J., PRICE T., VETTER T., BARTLETT J. C., BLANZ V.: 3d shape and 2d surface textures of human faces: the role of "averages" in attractiveness and age. *Image and Vision Computing* 18, 1 (1999), 9–19. 12
- [Pap06] PAPAGIANNAKIS G.: *An illumination registration model for dynamic virtual humans in mixed reality*. Dissertation, University of Geneva, 2006. 18
- [Pat95] PATEL M.: Colouration issues in computer generated facial animation. *Comp. Graphics Forum* 14, 2 (1995), 117–126. 12
- [Pel05] PELACHAUD C.: Multimodal expressive embodied conversational agents. In *Proceedings of the 13th annual ACM international conference on Multimedia, Brave New Topics session* (New York, NY, USA, 2005), ACM, pp. 683–689. 2, 9
- [PF02] PANDZIC I., FORCHHEIMER R.: *MPEG-4 Facial Animation: The Standard, Implementation and Applications*. John Wiley & Sons, West Sussex, England, 2002. 9
- [PG96] PERLIN K., GOLDBERG A.: Improv: a system for scripting interactive actors in virtual worlds. ACM, pp. 205–216. 5
- [PGL*06] PILET J., GEIGER A., LAGGER P., LEPETIT V., FUA P.: An all-in-one solution to geometric and photometric calibration. In *Int. Symposium on Mixed and Augmented Reality* (Santa Barbara, CA, 2006). 18
- [PGS08] PAN X., GILLIES M., SLATER M.: The impact of avatar blushing on the duration of interaction between a real and virtual person. In *Presence 2008* (2008), pp. 100–106. 12
- [PI04] PRENDINGER H., ISHIZUKA M.: *Life-Like Characters*. Springer, Heidelberg, 2004. 1, 4
- [PKL08] PETTRÉ J., KALLMANN M., LIN M. C.: Motion planning and autonomy for virtual humans. In *ACM SIGGRAPH 2008 classes* (New York, USA, 2008), ACM, pp. 1–31. 9
- [Plu80] PLUTCHIK R.: *Emotion: Theory, Research and Experience*. Academic Press, Sheffield, 1980. 11
- [PP04] PREDAM M., PRETEUX F.: Virtual character within mpeg-4 animation framework extension. *IEEE Transactions on Circuits and Systems for Video Technology* 14, 7 (2004), 975–988. 8
- [PPM*03] PONDER M., PAPAGIANNAKIS G., MOLET T., MAGNENAT-THALMANN N., THALMANN D.: VHD++ development framework: Towards extendible, component based VR/AR simulation engine featuring advanced virtual character technologies. *CGI '03: Proc. of CG Intl.* (2003), 96–106. 9, 18
- [PSI04] PRENDINGER H., SAEYOR S., ISHIZUKA M.: *MPML and SCREAM: Scripting the bodies and minds of life-like characters*. Springer, 2004, pp. 213–242. 7
- [PSKS04] PARK S. I., SHIN H. J., KIM T. H., SHIN S. Y.: On-line motion blending for real-time locomotion generation. *Comput. Animat. and Virtual Worlds* 15, 3-4 (2004), 125–138. 9
- [RCB98] ROSE C., CHOEN M., BODENHEIMER P.: Verbs and adverbs. multidimensional motion interpolation. In *IEEE CG and Applications* (1998). 9
- [RD00] REITER E., DALE R.: *Building Natural Language Generation Systems*. Studies in Natural Language Processing. Cambridge University Press, Cambridge, 2000. 4
- [Rey99] REYNOLDS C. W.: Steering behaviors for autonomous characters. In *Game Developers Conference 1999* (1999), pp. 763–782. 8
- [RH01] RAMAMOORTHY R., HANRAHAN P.: An efficient representation for irradiance environment maps. In *Proceedings of ACM SIGGRAPH 2001* (2001), Computer Graphics Proceedings, Annual Conference Series, pp. 497–500. 18
- [RJ99] RICKEL J., JOHNSON W. L.: Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence* 13 (1999), 343–382. 5, 6
- [Ros07] ROSADO G.: Motion blur as a post-processing effect. In *GPU Gems 3*. Addison-Wesley, 2007, ch. 27, pp. 575–581. 16
- [Rou00] ROUSSOU M.: Immersive interactive virtual reality and informal education. In *Proc. of User Interfaces for All: Interactive Learning Environments for Children* (Athens, 2000). 16
- [RSC87] REEVES W. T., SALESIN D. H., COOK R. L.: Rendering antialiased shadows with depth maps. In *SIGGRAPH '87* (NY, USA, 1987), ACM Press, pp. 283–291. 17
- [RSY03] RIEDL M., SARETTO C. J., YOUNG R. M.: Managing interaction between users and agents in a multi-agent storytelling environment. ACM, pp. 741–748. 5
- [SBH*90] SHEARN D., BERGMAN E., HILL K., ABEL A., HINDS L.: Facial coloration and temperature responses in blushing. *Psychophysiology* 27, 6 (1990), 687–693. 11
- [SCFRC01] SLOAN P.-P. J., CHARLES F. ROSE I., COHEN M. F.: Shape by example. In *I3D '01: Proceedings of the 2001 symposium on Interactive 3D graphics* (NY, USA, 2001), ACM, pp. 135–143. 9
- [Sch81] SCHANDRY R.: *Psychophysiologie: Körperliche Indikatoren menschlichen Verhaltens*. Urban & Schwarzenberg, München, Germany, 1981. 11
- [Sch84] SCHEGLOFF E. A.: On some gestures' relation to talk. In *Structures of Social Action*, Atkinson J. M., Heritage J., (Eds.). Cambridge University Press, 1984, pp. 266–296. 4
- [Sch04] SCHEUERMANN T.: Practical real-time hair rendering and shading. *Siggraph04 Sketches*, 2004. 10, 11
- [SGM04] SANDER P. V., GOSSELIN D., MITCHELL J. L.: Real-time skin rendering on graphics hardware. In *SIGGRAPH '04: ACM SIGGRAPH 2004 Sketches* (New York, NY, USA, 2004), ACM, p. 148. 13
- [SH07] STARCK J., HILTON A.: Surface capture for performance-based animation. *IEEE Comput. Graph. Appl.* 27, 3 (2007), 21–31. 8
- [SH09] SAFANOVA A., HODGINS J. K.: Synthesizing human motion from intuitive constraints. In *Artificial Intelligence Techniques for Computer Graphics*. Springer-Verlag, 2009, ch. 2, pp. 15–39. 9
- [SHG*01] SWARTOUT W., HILL R., GRATCH J., JOHNSON W. L., KYRIAKAKIS C., LABORE C., LINDHEIM R., MARSELLA S., MIRAGLIA D., MOORE B.: Toward the holodeck: Integrating graphics, sound, character and story. 6
- [SK08] STRAUSS M., KIPP M.: Eric: A generic rule-based framework for an affective embodied commentary agent. *Proc. of the 7th Int'l Conf. on Autonomous Agents and Multiagent Systems (AAMAS)* (2008). 6
- [SKG05] SUNG M., KOVAR L., GLEICHER M.: Fast and accurate goal-directed motion synthesis for crowds. In *SCA '05: Proceedings of the 2005 ACM SIGGRAPH/ Eurographics symposium on Computer animation* (New York, USA, 2005), ACM, pp. 291–300. 9
- [SKS02] SLOAN P.-P., KAUTZ J., SNYDER J.: Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. *ACM Transactions on Graphics* 21, 3 (2002), 527–536. 18

- [SOM*10] SWARZ J., OUSLEY A., MAGRO A., RIENZO M., BURNS D., LINDSEY A. M., WILBURN B., BOLCAR S.: Cancerspace: A simulation-based game for improving cancer-screening rates. *IEEE Computer Graphics & Applications* 30, 1 (2010), 90–94. 2
- [SS06] SUPAN P., STUPPACHER I.: Interactive image based lighting in augmented reality. In *Central European Seminar on Computer Graphics* (2006). 17, 18
- [ST04] SCHEUERMANN T., TATARCHUK N.: Advanced depth of field rendering. In *ShaderX3: Advanced Rendering Techniques in DirectX and OpenGL*. Charles River Media, 2004. 16
- [SWP10] SCHERZER D., WIMMER M., PURGATHOFER W.: A survey of real-time hard shadow mapping methods. In *EG 2010 - State of the Art Reports* (2010), Eurographics, pp. 21–36. 17
- [Tat06] TATARCHUK N.: Artist-directable real-time rain rendering in city environments. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Courses* (New York, USA, 2006), ACM, pp. 23–64. 14
- [TB08] TARIQ S., BAVOIL L.: Real time hair simulation and rendering on the gpu. *Siggraph '08 Sketches*, 2008. 10
- [TGB00] TOLANI D., GOSWAMI A., BADLER N. I.: Real-time inverse kinematics techniques for anthropomorphic limbs. *Graph. Models Image Process.* 62, 5 (2000), 353–388. 8
- [TM07] THALMANN D., MUSSE S. R.: *Crowd Simulation*. Springer, London, 2007. 2
- [TMMK08] THIEBAUX M., MARSHALL A., MARSELLA S., KALLMAN M.: Smartbody: Behavior realization for embodied conversational agents. In *Proc. of the Intl. Conf. on Autonomous Agents and Multiagent Systems* (2008). 4, 9
- [Tüm07] TÜMMLER J.: *Avatare in Echtzeitsimulationen*. Dissertation, Universität Kassel, 2007. 2, 7, 9, 10
- [VC01] VINGERHOETS A. J. J. M., CORNELIUS R. R.: *Adult Crying: A Biopsychosocial Approach*. Brunner-Routledge, San Francisco, 2001. 12
- [VCC*07] VILHJÁLMSOHN H., CANTELMO N., CASSELL J., E. CHAFAI N., KIPP M., KOPP S., MANCINI M., MARSELLA S., MARSHALL A. N., PELACHAUD C., RUTTKAY Z., THÓRISSON K. R., WELBERGEN H., WERF R. J.: The behavior markup language: Recent developments and challenges. In *IVA '07* (2007), pp. 99–111. 2, 5, 7
- [VGS*06] VINAYAGAMOORTHY V., GILLIES M., STEED A., TANGUY E., PAN X., LOSCOS C., SLATER M.: Building Expression into Virtual Characters. In *EG 2006 - State of the Art Reports* (2006), Eurographics Association, pp. 21–61. 2, 3, 11, 13
- [VMT04] VOLINO P., MAGNENAT-THALMANN N.: Animating complex hairstyles in real-time. In *VRST '04: Proc. of the ACM symposium on VR software a. technology* (NY, USA, 2004), ACM, pp. 41–48. 10
- [vTE09a] VAN TOL W., EGGES A.: 3d characters that are moved to tears. In *Short Paper and Poster Proceedings Computer Animation and Social Agents Conference* (Amsterdam, 2009). 14
- [vTE09b] VAN TOL W., EGGES A.: Real-time crying simulation. In *IVA 2009* (Heidelberg, 2009), Springer, pp. 215–228. 14
- [vWRRZ10] VAN WELBERGEN H., REIDSMA D., RUTTKAY Z., ZWIERS J.: Elckerlyc: A bml realizer for continuous, multimodal interaction with a virtual human. *Journal on Multimodal User Interfaces* (2010). 9
- [vWvBE*09] VAN WELBERGEN H., VAN BASTEN B. J. H., EGGES A., RUTTKAY Z., OVERMARS M. H.: Real Time Animation of Virtual Humans: A Trade-off Between Naturalness and Control. In *EG 2009 - State of the Art Reports* (2009), pp. 45–72. 3, 8, 9
- [Wah03] WAHLSTER W.: Towards symmetric multimodality: Fusion and fission of speech, gesture, and facial expression. In *Proceedings of the 26th German Conference on Artificial Intelligence* (2003), Springer, pp. 1–18. 4
- [Wah06] WAHLSTER W.: Dialogue systems go multimodal: The smartkom experience. In *SmartKom: Foundations of Multimodal Dialogue Systems*. Springer, Heidelberg, 2006, pp. 3–27. 2, 4
- [WB09] WEEKLEY J. D., BRUTZMAN D. P.: Beyond viewpoint: X3d camera nodes for digital cinematography. In *Web3D '09* (NY, USA, 2009), ACM, pp. 71–77. 16
- [WBJ69] WATZLAWICK P., BEAVIN J., JACKSON D.: *Menschliche Kommunikation. Formen, Störungen, Paradoxien*. Huber Verlag, Bern, 1969. 4
- [WBK*07] WARD K., BERTAILS F., KIM T.-Y., MARSCHNER S. R., CANI M.-P., LIN M. C.: A survey on hair modeling: Styling, simulation, and rendering. *IEEE Trans. Vis. Comput. Graph.* 13, 2 (2007), 213–234. 9, 10, 11
- [Web05] WEB3DCONSORTIUM: H-anim, 2005. <http://www.web3d.org/x3d/specifications/ISO-IEC-19774-HumanoidAnimation/>. 7
- [Web08] WEB3DCONSORTIUM: X3D, 2008. <http://www.web3d.org/x3d/specifications/>. 2, 7, 16
- [Wel93] WELMAN C.: *Inverse kinematics and geometric constraints for articulated figure manipulation*. Master's thesis, Simon Fraser University, 1993. 8
- [WL03] WARD K., LIN M. C.: Adaptive grouping and subdivision for simulating hair dynamics. In *PG '03: Proc. of the 11th Pacific Conference on CG and Applications* (2003), IEEE, pp. 234–242. 10
- [WLL*03] WARD K., LIN M. C., LEE J., FISHER S., MACRI D.: Modeling hair using level-of-detail representations. In *CASA '03: Proceedings of the 16th International Conference on Computer Animation and Social Agents (CASA 2003)* (2003), IEEE Computer Society, pp. 41–49. 10
- [WMT05] WANG H., MUCHA P. J., TURK G.: Water drops on surfaces. *ACM Trans. Graph.* 24, 3 (2005), 921–929. 15
- [YK08] YUKSEL C., KEYSER J.: Deep opacity maps. *Computer Graphics Forum (Proceedings of EUROGRAPHICS 2008)* 27, 2 (2008), 675–680. 10
- [YPWP05] YANG X., PETRIU D., WHALEN T., PETRIU E.: Hierarchical animation control of avatars in 3-d virtual environments. *IEEE Trans. on Inst. and Meas.* 54, 3 (2005), 1333–1341. 7
- [YT10] YUKSEL C., TARIQ S.: Advanced techniques in real-time hair rendering and simulation. In *SIGGRAPH '10: ACM SIGGRAPH 2010 Courses* (New York, NY, USA, 2010), ACM, pp. 1–168. 9, 10
- [Yu01] YU Y.: Modeling realistic virtual hairstyles. In *Proceedings of Pacific Graphics* (2001), pp. 295–304. 10
- [ZSXL06] ZHANG F., SUN H., XU L., LUN L. K.: Parallel-split shadow maps for large-scale virtual environments. In *VRCA '06: Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications* (New York, NY, USA, 2006), ACM, pp. 311–318. 17
- [ZYWK08] ZINKE A., YUKSEL C., WEBER A., KEYSER J.: Dual scattering approximation for fast multiple scattering in hair. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2008)* 27, 3 (2008), 1–10. 10