Virtual Humanoids and Presence in Virtual Environments

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der Rheinisch-Westfälischen Technischen Hochschule Aachen zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Diplom-Informatiker
Jakob T. Valvoda
aus Prag

Berichter: Universitätsprofessor Christian H. Bischof, Ph. D.
Universitätsprofessor Dr. rer. nat. Jan O. Borchers

Tag der mündlichen Prüfung: 14. September 2007

Abstract

The broad range of methods for virtual humanoids and their evaluation as avatars in virtual environments is handled in this thesis. The enabling methods, as well as experimental evaluations are the main contributions of the thesis. The central conclusion of the work can be summed up as the fact that virtual humanoids are beneficial for a number of interaction scenarios typically found in virtual environments.

The thesis presents the technology required for an interactive simulation of virtual humanoids in virtual environments. It presents the VRZula toolkit and its main features, mainly a layered data structure derived from functional anatomy, separation of data and simulation algorithms, and the control and propagation of modifications to the data. This elementary approach leads to an extendible and flexible system for virtual humanoids. The toolkit is capable of simulating even sophisticated models of virtual humanoids in complex distributed virtual environments. It integrates a variety of current approaches to the representation of virtual humanoids, as well as methods for kinematic and physiology simulation. Substantial improvements and novel approaches are introduced in the areas of deformation of musculature, multimodal representation and synchronization, construction of motion graphs, and synchronous body tracking.

Within the scope of this work, the VRZula toolkit is used for the evaluation of several hypotheses regarding the application of virtual humanoids in virtual environments. The two studies presented contribute to the understanding of the influence of virtual humanoids on the sensation of presence. It has been shown that avatars equipped with synchronous kinematics affect the real motion of the user and optimize the kinematics in the virtual world. Even though the mental and attentional resources bound to the exocentric virtual body representation led to a decrease of presence scores in room-mounted virtual environments, a strong identification of the user with the avatar has been shown. In addition, correlations between presence scores and physiological, as well as performance measures have been found. The results of the two studies are the basis of the evaluation of virtual humanoids used as avatars in typical interaction scenarios of virtual environments, and identify beneficial areas of application in certain navigation tasks and most manipulation scenarios. The use of avatars equipped with synchronous kinematics is clearly recommended.

The VRZula toolkit developed within this work and the experimental evaluations have contributed to the prevailing knowledge on virtual humanoids and their ap-

plication in virtual environments. The thesis demonstrates that virtual humanoids are the central methodology in future interaction techniques of sophisticated virtual environments.

Zusammenfassung

Das Kernthema dieser Arbeit behandelt die Simulation von virtuellen Humanoiden und die Evaluierung ihrer Nutzung als Avatare in virtuellen Umgebungen. Die Arbeit befasst sich unter anderem mit Simulationsmethoden, die den Einsatz von virtuellen Humanoiden in virtuellen Umgebungen ermöglichen. Den Hauptbestandteil allerdings bilden die experimentellen Untersuchungen der Zusammenhänge zwischen Benutzern und ihren Avataren. Die sich daraus ergebende Schlussfolgerung besagt, dass virtuelle Humanoide vorteilhaft in einer Vielzahl von typischen Interaktionsszenarien eingesetzt werden können.

Notwendige Technologien zur interaktiven Simulation von virtuellen Humanoiden in virtuellen Umgebungen werden in dieser Arbeit detailliert behandelt. Das hierfür entwickelte VRZula-System wird in seiner Gesamtheit beschrieben und die einzelnen Hauptmerkmale werden ausführlich behandelt. Insbesondere werden die Aspekte der mehrschichtigen, aus funktionaler Anatomie abgeleiteten Datenstruktur, die Kapselung der abstrakten Daten vor den Simulationsalgorithmen und die gezielte Steuerung und Regelung der Modifikation der Basisdaten beschrieben. Dieser strukturierte und grundlegende Ansatz führt zu einem System für virtuelle Humanoide, welches erweiterbar und flexibel in seiner Nutzung ist. So ist das System in der Lage, umfangreich modellierte virtuelle Humanoide in verteilten virtuellen Umgebungen zu simulieren. VRZula integriert nicht nur eine Vielzahl von Ansätzen zur Darstellung von virtuellen Humanoiden, sondern auch Strategien zur Simulation der Kinematik und der Physiologie. Grundlegende Verbesserungen werden in den Bereichen der Deformation der Muskulatur, der Synchronisation multimodaler Repräsentationsdaten, dem Aufbau von Bewegungsgraphen und dem Ganzkörper-Tracking präsentiert.

Das VRZula-System wird im Rahmen dieser Arbeit für die experimentelle Überprüfung von mehreren Hypothesen verwendet, welche den Einsatz von virtuellen Humanoiden in virtuellen Umgebungen hinterfragen. Es werden zwei zielgerichtete Studien vorgestellt, welche den Einfluss von virtuellen Humanoiden auf das Empfinden der Präsenz näher untersuchen. Die Studien haben gezeigt, dass mit synchroner Kinematik ausgestattete Avatare in der Lage sind, die Benutzerkinematik dermaßen zu beeinflussen, dass optimale Bewegungen im Virtuellen entstehen. Auch wenn die zusätzlichen, durch die Verwendung von exozentrischen Avataren induzierten mentalen Aufmerksamkeitsprozesse zu einer Verringerung des Präsenz-Empfindens in kopfferenen virtuellen Umgebungen führen, konnte die starke Iden-

tifikation der Benutzer mit ihren Avataren gezeigt werden. Darüber hinaus wurden Korrelationen zwischen Präsenz-Metriken und physiologischen Reaktionen ermittelt. Die beiden vorgestellten Studien bilden die Grundlage für eine ausführliche Diskussion der Verwendung virtueller Humanoide als Avatare in typischen Interaktionsszenarien. Navigationsaufgaben und insbesondere Manipulationsaufgaben werden in diesem Zusammenhang als vorteilhafte Nutzungsszenarien identifiziert. Eine Empfehlung zur Verwendung von Avataren mit synchroner Kinematik wird aufgrund der erarbeiteten Erkenntnisse ausgesprochen.

Das entwickelte VRZula-System und die vorgestellten Experimente bereichern den derzeitigen Kenntnisstand bezüglich virtueller Humanoide und unterstützen ausdrücklich ihre Anwendung in virtuellen Umgebungen. Die zentrale Rolle von virtuellen Humanoiden als die grundlegende Technologie für zukünftige Interaktionsmethoden in fortschrittlichen virtuellen Umgebungen wird durch diese Arbeit verdeutlicht.

Acknowledgements

This work has been carried out in the Virtual Reality Group at the Rechen- und Kommunikationszentrum of RWTH Aachen University from August 2001 to April 2007.

This work could not have been accomplished without the active support of my diploma and masters students Ingo Assenmacher, Sebastian Ullrich, Özenc Atca, Predrag Stojadinović, Stefan Korinek, Kalakota Ravisekhara Reddy, Joachim Börger, and Jens Eggerath to whom I wish to express my sincere appreciation.

I also wish to express my gratitude to Prof. Christian H. Bischof, Ph. D., for giving me the opportunity to carry out this work under his supervision and for many lucid, critical discussions and important suggestions. It is also a pleasure to thank Prof. Dr. Jan O. Borchers for co-reviewing this work.

My sincere appreciation goes to all the members of the Virtual Reality Group and its head Dr. Torsten Kuhlen, who in one way or another have supported this work. I would like to thank, in particular, Ingo Assenmacher, Marc Schirski, and Sebastian Ullrich, who always provided patient, thorough and understanding help. Many thanks go also to the Rechen- and Kommunikationszentrum collectively for the opportunity to use all of its facilities. This work was greatly stimulated by discussions with Dr. Christian Dohle whose suggestions and comments were essential in the neuropsychological part of this work.

Finally, I would like to express my sincere thanks to my family for the most valuable support: to Magdalena for her continuous encouragement, to Frieder for more than just proof reading but influential and essential suggestions, and to Adele and Alice for tireless polishing of my English. Yet the most important gratitude goes to my wife Kathrin for her permanent help and for simply giving me a relaxing time and the most wonderful perspective in life.

Aachen, June 2007

Jakob T. Valvoda

Contents

1

Introduction – The Simulation and Evaluation of Virtual Humanoids in Virtual Environments

Virtual environments are widely regarded as the future interaction medium for high complexity tasks that are seldom solved by present desktop systems. Virtual environments facilitate and enable, amongst other things, the exploration of large data sets difficult to understand in their complexity, the presentation and steering of complicated working procedures, or life-like and intuitive communication between distant parties.

Notwithstanding numerous attempts to improve the available techniques, serious deficiencies still exist with regard to the interaction and acceptance of the content presented in such environments. Unresolved issues are found, among others, in the areas of simulation of assembly processes, manipulation of virtual objects, navigation in virtual worlds, or communication in distributed virtual environments. These application scenarios typically challenge current development in terms of the occlusion problem during manipulation in room-mounted virtual environments, collision avoidance during navigation, or the general identification of the user with the virtual world. Methods for virtual environments thus continue to evolve.

Virtual humanoids have been broadly discussed as a solution to the prevailing problems by creating a more natural and life-like working experience and a frame of identification for the user. Not surprisingly, the development of methods for the simulation and assessment of virtual humanoids in virtual environments has become a key area of research. Research on virtual humanoids addresses the problem to create more or less comprehensive models of human beings and to simulate their

appropriate behavior. The application of such virtual humanoids for interaction purposes in virtual environments is expected to generate a more natural and life-like working experience. However, the impact of virtual humanoids on users and their interrelations during interaction have not yet been clarified to a satisfying extent.

This thesis focuses on virtual humanoids and their application in virtual environments. It begins with a detailed evaluation of present research in the field, with a particular emphasis on virtual humanoid toolkits. It goes on to outline simulation methods in use and continues with an overview of the current discussion on usability and presence focused on virtual humanoids. In this way, the subject matter of this thesis is introduced in the right perspective. After a brief exposition of the specific problems to be addressed (chapter 2), the results are set forth in two chapters:

- Chapter 3 addresses the development of new technologies for interactive virtual humanoids in virtual environments and describes the integration of current methods, flexible data structures, and novel real-time capable simulation techniques in the areas of representation, tracking, physiology and motion synthesis.

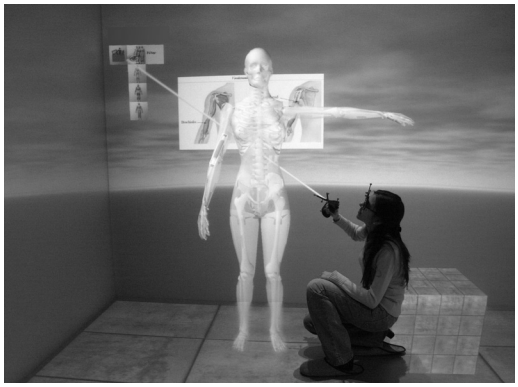
- Chapter 4 examines the influence of avatars, i. e., virtual user representations, on the sensation of presence in virtual environments and elaborates on the major influencing factors involved in this process.

These chapters are followed by an extensive discussion of the results achieved. In summary, the thesis is divided into a technical part (section 1.1 and chapter 3) and a part detailing the experiments and their theoretical foundation (section 1.3 and chapter 4). These parts can be read independently, even though the realization of the virtual humanoid toolkit is a *condicio sine qua non* for the studies undertaken.

This thesis brings forth several key issues thereby contributing to the research area of application, as well as to the evaluation of virtual humanoids in virtual environments, in a major way:

- The current approaches spanning the whole process of simulation of virtual humanoids are reviewed and presented in a capacious yet concise way. Several areas, like the simulation of kinematics or visual representation, are handled and explained in greater detail to allow for a coherent derivation of the novel methods proposed in this work.

- This thesis clearly focuses on the application of virtual humanoids in virtual environments. Virtual environments, in general, require interactive, immersive and multimodal representations of their content. This thesis presents the VRZula toolkit, which aims at a specific usage of virtual humanoids in virtual environments. The VRZula toolkit was designed to facilitate the creation of


a


b

Figure 1.1. Examples of application of virtual humanoids in virtual environments: (*a*) the simulation of physiology in a multi-layered humanoid model, and (*b*) a user observing her avatar in a room-mounted virtual environment. Virtual humanoids in both examples were created and simulated using the VRZula toolkit.

models of virtual humanoids that are based on structures derived from functional anatomy (*cf.* Fig. 1.1 *a*). The complexity of models is handled by the separation of data and simulation processes and through the distinct control and propagation of data modifications. The proposed framework allows for an interactive representation of complex virtual humanoid models and facilitates its application in distributed virtual environments.

- This thesis introduces several novel simulation approaches. These include the interactive, volume-preserving simulation of musculature with realistic appearance based on fast bounding-box computations; strategies for motion graph construction that allow for motion synthesis with improved realism and navigation capabilities; the synchronization of facial visual representation and auditory feedback of the spoken language; and a robust approach to orientation-based tracking. These methods were integrated into the VRZula toolkit and extend today's state-of-the-art methods for the simulation of virtual humanoids in virtual environments.

The influence of virtual humanoids on users was examined in two studies carried out as part of this thesis:

- The first study focused on the perception of self-kinematics mediated by a virtual avatar. This topic has been widely discussed in literature and linked to the concept of body schema in theory. Yet no experimental evaluation of cerebral activations in this context has been done so far. In this study, the motion of the user's arm was tracked and mapped onto the kinematics of an avatar. The mapping itself and the laterality of the displayed arm were altered. Functional imaging revealed cerebral activations in the precuneus that were strictly contralateral to the presented virtual arm. This indicated the

importance of synchronous kinematics for the acceptance of an avatar's body as one's own. The results of a kinematics analysis revealed one important fact: participants tended to adapt their own movements in a way that did not optimize the energetic effort in the motion of their own body but in the perceived motion of the virtual avatar. These findings emphasize the potential in applying avatars in virtual environments, in particular, their influence on the behavior of the user and on interaction metrics, and have motivated further experimental evaluations.

- The second study focused on the user acceptance of avatars with synchronous kinematics in room-mounted virtual environments. Besides performance metrics, the study used subjective and objective presence measurements to estimate the user preference in the virtual environment. The room-mounted set-up implied that users perceived their own body and the body of the avatar at the same time (*cf.* Fig. 1.1 *b*). The effects of avatars in room-mounted virtual environments on the sensation of presence have not been studied extensively to date. In this experiment, the subjects had to accomplish a sports task that required moving their entire body and stimulated their sense of vection and equilibrium. The kinematics of the avatar presented to the user were changed under three conditions and compared to a condition with no avatar. Presence questionnaires confirmed the hypothesis that subjects will feel the strongest presence, when no additional avatar is presented. Nevertheless, the physiological and performance measures acquired in this study allow a diversified interpretation of the interrelations between users and their avatars.

Virtual humanoids are an inevitable part of forward-looking applications in virtual environments. The reasons for this are manifold while most effects are still not well understood. Yet the strong identification with the representation of the virtual body and the intense induction of attentional processes outlined in this work demonstrate the capabilities of virtual humanoids. The use of virtual humanoids facilitates the creation of compelling virtual environments where content can be perceived in a more natural and realistic way. This thesis puts forward guidelines that help determine the suitable scenarios for applying virtual humanoids in virtual environments.

1.1 Present Technologies for Virtual Humanoids

Virtual humanoids have been discussed in the literature for more than four decades, with sources in environmental simulation, motion analysis, kinematic description, ergonomics and entertainment [BPW93]. Recent research mainly focuses on interactive application in human-computer interfaces [Tha00], and on hyper-realistic but non-interactive rendering in the film and entertainment industry, e. g., [Fer03, ST04].

Virtual Reality (VR) applications require interactive approaches to virtual humanoids that will be regarded in subsequent sections. According to [BPB99] several application areas for interactive virtual humanoids can be identified, for example, engineering, monitoring, and maintenance [BEL02], as well as virtual conferencing, games, training, and education [RJ99]. Existing solutions for these tasks, as well as current approaches to simulation techniques are presented in the following sections.

1.1.1 Toolkits for Virtual Humanoids

Numerous research activities focusing on virtual humanoid data structure design and simulation techniques have been reported. Interactive virtual humanoids are described in a comprehensive way in [BPW93] or [PI04]. Current toolkits for virtual humanoids are detailed in following paragraphs.

Jack. The *Jack* system [PB88] and the subsequent work TransomJack and JackMOO originally evolved from ergonomics studies. Jack is one of the first overall approaches to humanoid toolkits [BPW93]. This system contains an extensive skeletal model with anthropometric measures and algorithms for spatial interaction and visualization. Jack is suitable for the simulation of complex behaviors, such as grasping and walking. The application focus of Jack is the demonstration of various human capabilities and constraints, however, it can be applied in all scenarios that require a substitution of a real human in a workplace or other environment. Developed at the University of Pennsylvania, Jack is now a commercial product.

Critter. *Critter* is an animation toolkit whose main goal is to provide an efficient and effective animation environment for constructing and animating characters [CHP89]. The system uses layers—skeleton, soft tissue and musculature, and skin—and a motion control unit (also described as a layer). Kinematics and direct skin visualization can be done in an interactive way, however, the performance of the deformation of skin through musculature is described as "pseudo real-time" only. The focus of Critter is not on immersive virtual environments, yet it is one of the foundations for the layer concept in models for virtual humanoids.

DI-Guy. The *DI-Guy* (Dismounted Infantry Guy) is a product of Boston Dynamics[1] that started in 1992 as a spin off from the MIT Artificial Intelligence Laboratory. It animates interactive characters in real-time based on simple commands. The commercial system is used extensively in military applications.

LEMAN. Similar to the Critter system described above, the application of *LEMAN* is mainly in the area of creation of animations of virtual characters [TT93]. The system features a layered model consisting of a skeleton, musculature, soft tissue, and skin. The authors define interrelations of these components as a layered elastic model. The tool allows for a definition of keyframes and the final rendering of the animated sequence. Once again the focus of this system lies on designing and modeling animations instead of interactive application of virtual humanoids.

HUMANOID. Boulic *et al.* describe the *HUMANOID* environment [BCH+95] A further developed version is mentioned in [MT96]. This system supports the creation of articulated figures and emphasizes realistic deformation of skin. Particular attention is being paid to deformation of hands and facial expressions. The multi-layered model for simulation of human hands is described in [KMM+98]. Another topic is collision handling and optimization of computational tasks.

Improv. The *Improv* system constitutes an interactive tool for the creation of animated virtual actors [PG96]. It includes an animation engine that enables visualization of the virtual humanoid, diverse morphing and skin deformation algorithms, as well as kinematic simulation. The second component of Improv is a script-based behavioral engine. The system can be distributed in a network with users at remote locations. The authors report multiple applications of Improv in the areas of educational and shared virtual environments, as well as in interactive storytelling. Originally started at the New York University, the system has been developed by Improv Technologies since 1999, which suspended operations in 2002.

Steve. The focus of *Steve* (Soar Training Expert for Virtual Environments) is mainly guided teaching and tutoring for procedural tasks [JR97, RJ99]. The Steve system provides a virtual agent and pays special attention to perception and cognition of this virtual humanoid. The virtual tutor monitors user action within the virtual world and is able to demonstrate movements and to advise the user in case of failure or errors in the procedure.

Lokutor. A further agent system is described in [Mil00]. The *Lokutor* system represents an animated agent that is meant to be autonomous from other

[1] http://www.bostondynamics.com, last visit May, 2007.

system components and to behave in a meaningful way. Lokutor consists of a behavioral and a motivational system and uses an underlying knowledge base of the virtual world. The system also integrates a speech recognition module that enables reactions to spoken commands. The author reports an application of the system for the presentation of virtual objects.

SimHuman. The aim of the *SimHuman* system is to enable intelligent virtual agents in desktop-VR applications [VP01]. The system has an embedded physics engine and supports path planning and kinematics, as well as goal-oriented behavior, which is realized by a Prolog-based solver.

BEAT. A system with focus on speech interaction is *BEAT* [CVB01]. It accepts text as input and produces synthesized speech, as well as synchronized non-verbal behavior. This data can be passed on to animation systems for display and animation of the virtual humanoid. The system is described as a helper tool for the creation of animations.

VHD++. The *VHD++* framework is a general toolkit for virtual and augmented reality [PPM+03]. The framework is addressing a broad range of applications. In particular, virtual humanoids can be included as a service component. VHD++ humanoids contain skeletal animation and kinematics, skin and facial deformation and cloth simulation. A voice component can be added, as well. The framework supports the simulation of multiple humanoids and human crowds. Applications of the framework and its humanoid service are described in the areas of VR-based training for rescue situations, maintenance, and visualization of historical sites.

VHI. An approach that integrates direct user interaction and realistic virtual humanoids is described in [TK03] and named *VHI*. This system has the ability to react on user interaction derived from facial feature extraction. VHI includes techniques for kinematics and focuses on emotion expression through facial animation. Interactive frame rates are reported, as well as the use of VHI with auto-stereoscopic devices.

Virtual Human. Interactive virtual humanoids are the main focus of the *Virtual Human* project [GSI+04]. The system consists of multiple components comprising modeling and authoring tools, storytelling and dialog, as well as visualization and kinematic simulation modules. Main application areas are learning scenarios in virtual environments. The narrative part is described in a markup language. Adapters to internet-based viewers are also discussed.

VHA. A recent contribution describes the *VHA* (Virtual Human Architecture) system [IC05]. H-Anim data is taken as basis for humanoid structure definition. The main focus of this tool is the simplification of definitions of behavior for

virtual humanoids. An extension to the H-Anim standard is being proposed covering structural and kinematic descriptions on higher levels.

1.1.2 Classification of Simulation Methods

The human body is a complex organism with a number of processes taking place on different scales: from the microscopic, cellular range and neuronal processing up to macroscopic actions of the whole body. In order to handle this complexity current approaches often focus on selected areas and utilize manifold techniques. The general approach in designing and simulating virtual objects implies definition and interrelation of form, function, and behavior (e. g., [KKKL98, LKK00]). This concept is extended and further specified by Funge [FTT99] who classified the following levels of simulation in a computer graphics modeling hierarchy: geometry, kinematics, physics, behavior and cognition. The algorithms presented in the following sections are roughly grouped according to Funge. The focus are techniques for visualization, kinematic simulation, motion blending, and musculature simulation.

1.1.3 Visualization Methods

Visualization of virtual humanoids is closely related to methods proposed in the areas of computer graphics and visualization. In general, a virtual humanoid is represented as one or multiple triangular meshes and textures that can be directly visualized. The use of voxels and implicit volume-rendering can be found in medical applications but such approaches are not wide-spread for interactive virtual humanoids. Visualization of humans with point-based rendering is reported in some works, e. g., by Gross *et al.* [GWN+03]. This technique is useful for interactive visualizations of densely sampled surfaces but the visual quality may suffer for data with a coarse resolution as it is mainly used in the display of virtual humanoids. General visualization techniques range from non-interactive offline rendering processes that achieve highly realistic results to real-time visualizations that often impose a limited realism. During the visualization process the geometrical representation is deformed and adapted to current parameters, e. g., facial expressions and body posture derived from kinematics, physiology, or behavior simulations. Afterwards, the adapted representations are rendered appropriately.

A humanoid model is typically changing its state continuously. Several techniques have used and interpolated sequences of adapted representations in order to create animated visualizations. This quite complicated process and the frequent use of kinematic simulation has motivated additional visualization methods relying mainly on posture data. These general approaches to real-time visualization techniques for virtual humanoids are presented in subsequent sections.

1.1.3.1 Interpolation of Models

The first techniques for visualization of virtual humanoids have been adopted from animated films. The animator creates a sequence of representations of the virtual humanoid. Each representation is associated with a particular time value, the so-called *keyframe*. The current representation G is computed from adjacent keyframes depending on the animation time t. Let $K_1, \ldots, K_n$ be the involved keyframe representations, and $f_i(t)$ a weighting function dependent on time and the keyframe. The time-dependent representation is computed as $G(t) = \sum_i f_i(t) \cdot K_i$. Even though this approach allows for a direct control of the visual representation it is limited by the fact that the sequence of keyframes determines the entire visualization. Furthermore, the creation of keyframe sequences requires significant time and resources.

The original key framing idea is extended in the *blendshape* method, also known as morph target or shape interpolation method. In this approach, the particular models K_i are not dependent on time or sequence restrictions but directly associated with parameters. Multiple geometries, so-called morph targets, are blended with a predefined mapping. Let the previously mentioned $K_1, \ldots, K_n$ be the morph targets. The current representation G is computed as $G = \sum_i w_i \cdot K_i$, where w_i represents the weight of parameters associated with morph target K_i. This method is often applied for visualizations of expressions of the face. Parke uses this technique to visualize emotion expressions [Par72] and suggests to use a cosine function for interpolation of the morph targets. Further applications of the blendshape method are described in [Kle89, PHL+98, JTDP03, LMDN05, DCFN06]. In these approaches blendshapes are used to represent visemes for speech rendering (*cf.* section 1.1.4) and for basic emotional expressions. The blendshape technique is suitable in the case that particular representations can be associated with parameters, like emotion expressions or mouth movements for speech representation. Even though the method can deliver results with a high degree of realism it is limited in terms of visualization of changing postures.

1.1.3.2 Posture-based Methods

Kinematic simulation is often applied in combination with virtual humanoids. Blendshape methods, as described in the previous section, are limited in their use for representation of changing postures. The basic visualization technique for pose representation is the *rigid skin* approach in which the humanoid is represented by multiple geometries $G_i = (V_i, T_i)$, each one consisting of a set of vertices V_i and connecting triangles T_i defined within the local coordinate system of the underlying kinematic segment. The current spatial transformation of each geometry $\tilde{G}_i$ is directly computed with the global transformation $\tilde{M}_i$ of the adjoining segment, resulting in $\tilde{G}_i = (\tilde{M}_i \cdot V_i, T_i)$. In this approach geometric primitives can be used, e. g., for the visualization of the kinematic structure and more detailed and elaborated models for representation of an anatomical skeleton, an artist's mannequin, or a seg-

mented visualization. Obviously, this approach allows for a limited degree of realism especially in the area of joints, but has a well optimized rendering performance.

The biggest drawbacks of the rigid skin visualization are possible gaps appearing at joint limits. These visual artifacts can be eliminated by connection of the rigid parts through a patch. This technique is referred to as *stitching* [Woo00]. Bordering rigid geometries are searched for edge contours and connected with a triangle strip. The vertices of the strips are transformed according to the position and orientation of the rigid geometries. This technique has a similar degree of realism as the rigid skin technique but ameliorates the approach by eliminating the coarsest visual artifacts.

Beside these basic techniques the *vertex blending* method is often used for visualization of virtual humanoids. As stated in [AH02, Chapter 3.4] and [Lan98b], origins of the technique are described in multiple sources. Alternative names are skinning [Woo00], enveloping [WP02], and skeleton subspace deformation [LCF00]. In this approach, the humanoid is represented as one vertex mesh, i. e., one geometry $G = (V, T)$. Each vertex $\mathbf{v}_j \in V$ is associated to one or multiple segments of the kinematic simulation with specific weights $w_{i,j}$. The transformed vertex $\mathbf{v}'_j$ is computed through a weighted sum of transformations of all adjoining kinematic segments $\mathbf{v}'_j = \sum_i w_{i,j} \cdot \tilde{M}_i \cdot (\tilde{M}_i^0)^{-1} \cdot \mathbf{v}_j$ where $\tilde{M}_i$ represents the global transformation of the associated segment i, and $\tilde{M}_i^0$ its initial transformation, the so-called dress-pose. This technique is used in most applications of interactive virtual humanoids because it allows for a smooth representation of the humanoid and can be computed in an efficient way[2]. However, the visual quality of the resulting deformation often resembles the bending of plastic tubes instead of human skin.

Because vertex blending uses a linear combination of transformations in order to compute the final vertex position it exhibits some typical problems. The most prominent ones are often referred to as "vertex collapse" and "candy wrapper" [LCF00], which occur on strongly bent or twisted joints. Vertex blending is able to visualize the humanoid by means of a smooth geometry well adapted to the underlying kinematic structure, yet it does not take into account further motion and deformation of the natural body, e. g., breathing movements or muscle bulges. In their pose space deformation approach Lewis *et al.* combine vertex blending with interpolation techniques in order to achieve smoother blending without the previously mentioned problems [LCF00]. However, this approach requires that natural deformations of the geometry have to be specified for particular poses. In similar approaches the linear blending system is combined with additional factors that are computed from a set of examples containing skeleton data for varying poses and accordingly deformed geometries. A recent generalization of the method is presented in [MMG06]. Wang and Phillips use a matrix of weights instead of one scalar value to compute the linear combination. The weights are determined and derived from a set of examples [WP02]. Sloan *et al.* present a method for interpolation between multiple examples [SRC01]. In a further approach Kry *et al.* derive pose-dependent

[2]Recent graphics hardware allows for a computation directly on the GPU.

displacers from the set of examples [KJP02], using an eigenvalue analysis in order to determine the essential displacers for each joint. Allen *et al.* use examples of a human torso [ACP02]. They refit a surface template and a skeleton to each example and use it for the computation of displacements. The final skin representation is interpolated based on the current pose and the particular displacers. Mohr and Gleicher use a set of examples to determine additional joints for the skeletal structure for alleviation of the linear errors of the basic vertex blending approach [MG03]. The example-based methods presented are capable of ameliorating the basic vertex blending approach. However, they assume that a set of examples and often the corresponding skeletal configurations are given—a cumbersome process for animators. This is also supported by the fact that most of these approaches are presented with selected regions, only, e. g., hands, torso, and arms. A solution for the entire body has, as yet, not been advanced.

Still another group of skeleton-driven simulation approaches uses general deformation techniques, like free form deformation (FFD) [SP86] or finite element (FE) methods. In general, deformations are determined through simplified objects, control lattices or points, linked to the kinematic structure of the skeleton. FFD is used in the work by Singh and Kokkevis [SK00] implying a control lattice derived from and oriented on the skin surface. The influence of the control lattice on each vertex is calculated using a distance metric. After manipulation of the controller the new vertex is computed as a weighted sum. This approach is comparable to the vertex blending approach in that it linearly blends multiple influencing factors. The point of this approach is that a more complex control structure is used resulting in additional summands in the weighted sum. However, the deformation of the control lattice itself becomes less practicable. Furthermore, it is difficult to bind the deformer to the skeletal structure to obtain skeleton-based results. In a more recent approach, Capell *et al.* use FEs to deform virtual characters [CGC+02]: a hull constructed from FEs encloses the character volume and is connected to the underlying skeleton. Changes in the configuration of the skeleton impose an FE simulation, which results in a deformation of the skin surface. According to the authors, this technique is not focused on anatomical modeling but tries to deform the volume of the solid object irrespective of its internal structures. A similar approach is used by Guo and Wong who apply deformable chunks for skinning [GW05]. Hyun *et al.* use a control structure that is automatically determined from the original skin surface and derive deformations from modifications on these structures [HYC+05]. These deformation approaches extend the original ideas of vertex blending in that more complex control structures and volumetric simulations are used, but this advantage is leveled through difficult manipulations of the control structure.

Posture-based techniques are widely spread in applications containing interactive virtual humanoids. The most important advantage is the direct connection to the kinematic structure that allows for animated or easy control. Drawbacks are a limited degree of visual realism resembling deformations of artificial objects. Although example-based methods and general deformation approaches have increased

the visual quality, they require a particularly large set of examples or impose a more complex control of the deformation algorithm. In order to increase the visual degree of realism several techniques have tried to simulate internal structures and to deform the skin surface according to these simulation results. These approaches are described in the next section.

1.1.3.3 Deformations Based on Internal Structures

It is obvious that the best method for creating a realistic visualization would be to simulate the anatomical, physiological, and biomechanical processes of a human body and to visualize the results. Some deformations, e. g., muscle bulges, skin wrinkles, or breathing motions of the thorax, would directly affect the skin representation and yield a high degree of visual realism. Unfortunately, this approach is still not applicable for interactive virtual humanoids because of its complexity and the high computational costs. The interpolation methods described previously are capable of visualizing some anatomical deformations but at the cost of time intensive creation of additional keyframe models or blend targets. Posture-based methods try to address this problem with additional artificial kinematic segments or deformations that have been derived from a number of examples. The results are compelling but limited to deformations and regions given by the examples, and, therefore, are not easily adaptable to additional structures. Lastly, several methods have been presented that approximate the overall body as deformable tissue. Limitations of these methods are mainly due to this general simplification. Despite the previously mentioned performance issues for a more detailed simulation of internal structures, some methods have been reported that try to simulate the anatomy of the body in order to produce more realistic simulation results, as described in the following.

Chadwick *et al.* propose to use a muscle and fatty tissue layer, *cf.* [CHP89] and section 1.1.1. This soft tissue is approximated by parametric solids whose control points are linked to a skeleton and used in an FFD approach to deform the skin representation. The approach by Turner and Thalmann uses a skeleton, a musculature, and a skin layer [TT93]. Musculature is dependent on the underlying skeletal structure and approximated by spheres and implicit surfaces. These deformable objects affect a fat layer, which defines an offset to the final skin layer. The authors suggest to represent the skin as a mesh of points connected with strings. A technique using a subclass of implicit surfaces, so-called metaballs, for the representation of internal structures is described in [ST95]. Metaballs are combined with B-splines and connected to a skeletal structure for simulation purposes. This approach is incorporated in a system called Body Builder [TSC96]. The rendering of implicit surfaces is, with respect to interactive virtual humanoids, the biggest drawback of this approach: it does not currently allow for interactive frame rates. For applications in virtual environments the authors propose to triangulate the implicit surface in the dresspose and to deform it with methods described in the last section. It is not surprising, that this solution does not provide advantages of a simulation of internal structures.

Wrinkles are another example for visualization based on biomechanical processes and internal structures. Wu *et al.* describe a model for simulation of wrinkles [WKM96]. They utilize musculature for expressive wrinkles. The skin is defined according to a biomechanical model. Examples for hand wrinkles and visualization of facial expressions are given. An early approach to the simulation of the human face is described in [LTW95]. In an early work Gourret *et al.* define an FE model in order to simulate the deformations of a hand during grasping [GMT89]. This approach requires about one minute to compute a simulation frame. In a more recent approach Teran *et al.* use FEs as well to model musculature used for skin deformation [TSIF05]. Kähler *et al.* describe a system for reconstruction of the face using a musculature layer [KHS03]. A comparable technique is chosen by Pratscher and colleagues who describe a system that is based on a skeleton and a skin model [PCLS05]. Muscles are modeled as ellipsoids and subsequently reshaped in that they fill the gaps between skeleton and skin. Muscles are deformed following the approach described in [SPCM97] by explicitly computing and adjusting the volume of the final ellipsoids. In addition, concurrent muscle definitions can be used for different poses and blended according to the approach described in [LCF00]. Finally, the musculature shapes are used for deformation of the skin surface. These approaches demonstrate on the one hand the use of finite elements and, on the other hand, the application of musculature for deformation of the skin. A realistic simulation of the deformation of soft tissue can be achieved with finite elements, however, the methods often do not allow for interactive frame rates. In order to achieve real-time processing the simulation models have to be simplified or focused on selected regions, which reduces the advantages over techniques that only approximate the internal structures, as described, for example, in the previously mentioned paper by Pratscher and colleagues. The simulation and application of musculature is a promising approach, however no visually sophisticated, real-time capable methods have been presented so far. Techniques to musculature simulation are described in greater detail in section 1.1.6.

Visualizations based on deformations derived from anatomical structures enable simulations with a high degree of visual realism. The presented methods use simplified or abstract models for internal structures. This simplification imposes some limitations on the overall simulation procedure. So far, no sophisticated solution that addresses both degree of realism and interactivity has been presented. Methods that use a more detailed simulation of musculature (see section 1.1.6) are capable of creating results with a better realism of appearance, however, the focus of these methods is mainly on visualization of internal structures instead of visualization of the skin. To conclude, further research on real-time visualization methods for virtual humanoids with an increased degree of realism has still to be done.

In current interactive approaches, mainly vertex blending is utilized for visualization of virtual humanoids. An increase in the degree of realism has mainly been achieved through high-resolution textures and the use of specific shaders for the rendering process. However, these techniques are limited and the natural, lo-

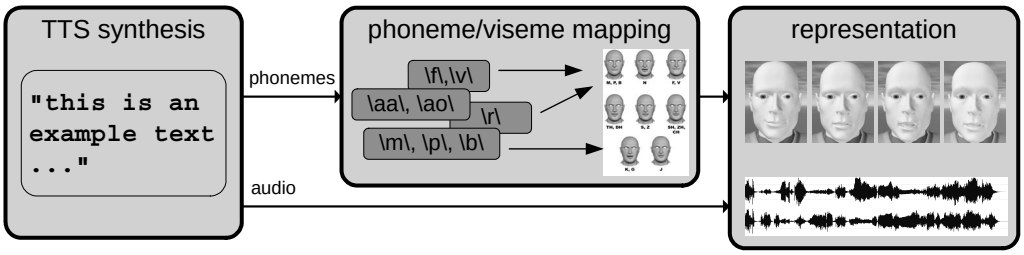


Figure 1.2. Processes involved in speech synthesis and audio-visual representation.

cal deformation on the human skin, e. g., through muscle bulges or breathing, can not be modeled with this approach. Example-based techniques seem to be promising to reach a high degree of realism but have the disadvantage that a partially large amount of examples has to be generated in order to train the system and to determine pose-dependent local deformations. Similarly, the blendshape technique achieves a sophisticated degree of realism in appearance if highly realistic geometry targets are used, but is also completely limited to these models. The research of visualization methods that include underlying structures of the human anatomy and their simulations has not resulted in satisfactory methods, so far.

1.1.4 Synthesis and Representation of Speech

In search of intuitive and natural ways of communicating with interactive systems, multimodal interfaces like embodied conversational agents (ECA) seem to be the designated successors to the currently dominating point-and-click interaction metaphor. Since the most natural form of interaction is a human face-to-face conversation, communicating with a virtual character is the most promising way of controlling interactive systems. According to this notion an integral characteristic of a conversational agent is its ability to speak. The general topic of audio-visual simulation of speech for virtual environments involves several areas of speech synthesis, their synchronization, and methods for immersive representation of multimodal data.

Spatial audio rendering is the auditory counterpart to stereoscopic visualization. The listener perceives auditory events originating from a certain location in space different from its real physical source. Available techniques are categorized in binaural synthesis and multi-channel auditory displays. Binaural synthesis utilizes the fact that signals emitted from a sound source hit the ears with a difference in time and in amplitude level. The brain uses this information to localize the source of the signals. Binaural synthesis artificially creates individual signals for each ear, simulating those natural conditions. Multi-channel auditory displays reproduce a spatial sound field by using an array of loudspeakers, so a rather complex hardware set-up is needed. Furthermore, most techniques produce correct signals for exactly one spot, thus, the user is not allowed to move. Considering these general methods

the binaural synthesis approach is more flexible. It allows for the use of two or four loudspeaker only and can be combined with head tracking.

Subsequent sections present current methods for synthesis of speech and integrated systems for audio-visual representation.

1.1.4.1 Synthesis of Speech

The most common method for creation of artificial speech is to use and process text as input data. This approach is usually referred to as *text-to-speech* (TTS) synthesis. The synthesis process consists of two major steps: the natural language processing (NLP) and the digital sound processing (DSP) [Dut97]. NLP maps the input text to a sequence of phonemes, the so-called phonetic transcription. Phonemes are considered as the smallest phonetic unit with regard to the meaning of an utterance. The phonetic transcription is extended with prosodic information, like accentuation and a basic frequency curve. The final DSP stage yields the wave representation of the audio data. A prominent freely available speech synthesis software is the *Festival* system developed by Taylor and colleagues [TBC98]. Festival is characterized by its modular and open structure. A commercial system is *RealSpeak* by Nuance Software[3]. The most striking difference compared to Festival is RealSpeak's high performance with regard to the synthesis process. A good overview on further currently available TTS systems can be found on the MBROLA project page[4].

1.1.4.2 Integrated Audio-visual Systems

Systems for production of audio-visual speech range from simple talking heads to intelligent ECAs. If plain text serves as input data, such kind of systems is referred to as *text-to-audio-visual-speech* (TTAVS) systems. The general process is illustrated in Fig. 1.2. *Baldi* is a prominent example for a TTAVS system, which evolved over more than a decade [MOCC05]. Originally created as a talking head for training of lip reading for the hearing impaired, Baldi's anatomical structure was gradually improved with captured data of real faces. Nowadays Baldi represents an intelligent ECA, acting as a language teacher. The system is designed for TTS software that is compliant with the Microsoft Speech API. The multimodal dialog system *SmartKom* enables a user to interact with a cartoon-like humanoid character on systems with touch sensitive PDAs and tablet PCs [PM02, RAB+03]. Göbel and colleagues describe the *Virtual Human* platform [GSI+04]. The system provides intelligent virtual characters as personal dialog partners for application in educational domains (*cf.* section 1.1.1). Cassell *et al.* present general design considerations concerning the conversational behavior of virtual characters and apply those to an ECA, which mimics the role of a virtual real estate agent [CVB01]. Most of the mentioned systems are characterized by a distributed, modular struc-

[3] http://www.nuance.com, last visit May, 2007.

[4] http://tcts.fpms.ac.be/synthesis/mbrola.html, last visit May, 2007.

ture to encapsulate several aspects of controlling and displaying a conversational agent. The presented systems integrate auditory data and visual information and generate the desired multimodal representation. However, immersive representation has not been handled extensively. In particular, none of the systems enable binaural sound rendering and stereoscopic visualization in an interactive virtual environment. In principle, the immersive methods for representation could be applied at the end of the processing pipeline, however, the systems typically do not allow for a modular access to the audio-visual data. In addition, methods for synchronization have to be reconsidered for immersive techniques. Another issue concerns the runtime behavior of the systems. Since ECAs are utilized in interactive scenarios, the responsiveness and the real-time capability have a high impact on the user acceptance. Beringer and colleagues develop an evaluation framework for the SmartKom project, but mainly focus on quality related values [BKL+02]. Pfleger and Löckelt propose to shorten long response times by providing additional feedback based on the input of the user, before the final response is calculated [PL05]. But like most contributions in the field of audio-visual speech and intelligent reactive systems, they do not discuss the reasons for such weak responsiveness in terms of the runtime performance of the individual components.

A crucial step in audio-visual speech is the mapping of visual and auditory data described in last section. The simplest approach is to find a corresponding facial expression (*viseme*) for each phoneme and to provide a transition from one viseme to the next depending on the chosen animation technique. The second aspect of synchronization is an exact chronological alignment of the representation of speech and facial animation. Asynchronisms are easily detected if the amount of audio or video delay exceeds a certain boundary, with video delays being more critical than delayed audio output [GvWP04]. For an even more natural behavior of a talking virtual humanoid, nonverbal actions like gestures have to be taken into consideration. Kopp and Wachsmuth address the problem of synchronizing speech and gestures [KW04]. To be able to synchronize speech with the corresponding visual output it is imperative to analyze the timing behavior of the overall system.

1.1.5 Motion Simulation

The simulation of motion and kinematics of articulated structures are central aspects for each interactive virtual humanoid. The kinematics of a human body is determined through several degrees of freedom (DOFs) of joints that are hierarchically represented in a skeletal structure. An extension to a purely hierarchical definition of the articulated structure is proposed in [SRCP02]. Usually, the DOFs are animated or selectively manipulated and result in a motion of the virtual humanoid. In the ensuing paragraphs the approaches towards the creation of desired postures or motions are described.

1.1.5.1 Direct Determination of Kinematics

The basic technique of *forward kinematics* is advantageous for specific pose creation through direct joint manipulation. The motion can be described in terms of the kinematic function $F(\mathbf{q})$, which for a joint results in its particular translation $\mathbf{v}$ and rotation q

$$F(\mathbf{q}) = [\mathbf{v}, q]^T \equiv M$$

where $\mathbf{q}$ is a vector of quaternions q_i representing the rotations of the kinematic chain. However, the direct manipulation of multiple rotations q_i can be a complex and difficult to handle task. In order to simplify this process *inverse kinematics* approaches have been developed. These methods accept a desired transformation M_d of a joint, the so-called end effector, and compute a rotational configuration of the kinematic chain $\mathbf{q}_d$ such that $F(\mathbf{q}_d) = M_d$ for the end effector. Therefore, these algorithms try to find the inverse representation $F^{-1}(M)$ of the kinematic function. Several approaches for the solutions of inverse kinematics have been reported and are described in following paragraphs.

Analytical Methods. The inverse kinematics problem can be solved directly, resulting in so-called closed-form solutions. In these approaches the kinematic equation is solved for the particular DOFs. It is obvious that such an approach is becoming increasingly complex for a greater number of DOFs. Early contributions propose a closed form solution for up to 9 DOFs [PSM81]. Manocha and Canny describe a solution for a 6 DOF manipulator [MC94]. They use the Denavit-Hartenberg notation [DH55] and an eigenvalue analysis in order to solve the kinematic problem. Lee and Shin present a closed-form solution for human limbs with 7 DOFs [LS99] and use quaternions for joint rotations. Tolani *et al.* propose another analytical solution for a human arm consisting of 7 DOFs [TGB00]. A recent publication by Wu *et al.* describes a solution of a kinematic chain of the human arm containing 12 DOFs [WMCG04].

The previously mentioned method by Lee and Shin uses the length of the upper arm l_u and the forearm l_f, as well as the rotational axis n_e defined by the shoulder's DOFs [LS99]. Given the distances r_u and r_f between the axis n_e and the shoulder and the wrist, respectively. The elbow rotation θ that results in the desired shoulder–wrist distance l_d is given as

$$\theta = \cos^{-1}\left(\frac{\lambda + 2\sqrt{\rho_u \cdot \rho_f} - l_d^2}{2 r_u r_f}\right)$$

with $\lambda = l_u^2 + l_f^2$, $\rho_u = l_u^2 - r_u^2$, and $\rho_f = l_f^2 - r_f^2$. With the elbow orientation adjusted to θ the shoulder joint can be rotated in that the wrist and the elbow are in the desired position. Finally, the 3 DOFs of the wrist are updated to achieve the desired orientation of the end effector.

Analytical methods are helpful for simulation of limbs or rather small parts of kinematic chains, but not applicable for whole humanoid structures, which consist

of considerably more joints. More general and flexible techniques are described in subsequent paragraphs.

Jacobian-based Methods. The majority of techniques for inverse kinematics are based on the Jacobian matrix $J = F'(\mathbf{q}) = \partial F(\mathbf{q})/\partial \mathbf{q}$, which is the multi-dimensional derivative of the kinematic function $F(\mathbf{q})$. The kinematic function can be rewritten as a Taylor series

$$F(\mathbf{q} + d\mathbf{q}) = F(\mathbf{q}) + F'(\mathbf{q}) \cdot d\mathbf{q} + \cdots + \frac{F^{(n)}(\mathbf{q})}{n!} \cdot d\mathbf{q}^n + \cdots .$$

By adopting the first term in the Taylor series and omitting the rest a linear approximation of the kinematic function is given as

$$F(\mathbf{q} + d\mathbf{q}) = F(\mathbf{q}) + F'(\mathbf{q}) \cdot d\mathbf{q} .$$

This equation can be reformulated yielding a description of the relation between joint rate $d\mathbf{q}$ and change in the transformation of the end effector dM

$$\begin{aligned} F(\mathbf{q} + d\mathbf{q}) - F(\mathbf{q}) &= F'(\mathbf{q}) \cdot d\mathbf{q} , \\ dM &= J \cdot d\mathbf{q} . \end{aligned}$$

As mentioned before, the goal in inverse kinematics is to compute the required change in the joint configuration $d\mathbf{q}$ such that a desired end effector transformation M_d is reached. By defining $dM = M_d - M$ the equation

$$d\mathbf{q} = J^{-1} \cdot dM$$

is yielding the wanted result. However, two problems occur in this reformulation: the computation of the Jacobian itself, and the definition of the inverse problem J^{-1}.

Computation of the Jacobian. The Jacobian represents the multi-dimensional derivative of the kinematic function $F(\mathbf{q}) = [\mathbf{v}, q]^T$ and is a matrix of partial derivatives

$$J = \frac{\partial F(\mathbf{q})}{\partial \mathbf{q}} = \left[\begin{array}{ccc} \frac{\partial F(\mathbf{q})}{\partial q_1} & \cdots & \frac{\partial F(\mathbf{q})}{\partial q_n} \end{array} \right] .$$

Each joint may have up to three rotational DOFs that can be specified as angular values around preferably linearly independent rotational axes or—in order to stay with the quaternion notation—as the logarithm of the particular quaternion $\log(q_i) = [\theta_i^x, \theta_i^y, \theta_i^z]^T$ (*cf.* [DKL98], and similar ideas of an exponential map representation for quaternions presented in [Gra98]). The partial derivatives of the Jacobian are, therefore, given as

$$\frac{\partial F(\mathbf{q})}{\partial q_i} = \left[\begin{array}{ccc} \frac{\partial \mathbf{v}}{\partial \theta_i^x} & \frac{\partial \mathbf{v}}{\partial \theta_i^y} & \frac{\partial \mathbf{v}}{\partial \theta_i^z} \\ \frac{\partial q}{\partial \theta_i^x} & \frac{\partial q}{\partial \theta_i^y} & \frac{\partial q}{\partial \theta_i^z} \end{array} \right] \in \mathbb{R}^{6\times 3} .$$

There are several ways to compute the Jacobian. Analytical methods and techniques for numerical differentiation can be applied. An approach for efficient computation of Jacobian matrices for varying arrangements of a kinematic chain is presented in [OS84]. Other methods are derived from geometrical observations. Another method is proposed by Whitney [Whi72]. Let each joint i have a distinguished rotational axis $\mathbf{u}_i$, the current position of the joint $\mathbf{v}_i$, and one rotational DOF around $\mathbf{u}_i$ given as the angle ξ_i. The columns of the Jacobian are given as

$$\frac{\partial F(\mathbf{q})}{\partial \xi_i} = \begin{bmatrix} \mathbf{u}_i \times (\mathbf{v} - \mathbf{v}_i) \\ \mathbf{u}_i \end{bmatrix} \in \mathbb{R}^6 .$$

A prominent approach for computation of the Jacobian without truncation error is automatic differentiation (AD) [Gri00]. AD tools have been developed for a number of languages, like FORTRAN [BCKM96], or C [BRM97], and are capable of computing derivatives even for large computational systems [BBR+07].

Solutions for the Inverse Problem. With a given Jacobian matrix the task is to formulate the inverse problem. The Jacobian is in the majority of cases not square[5]. Therefore, a simple inverse can not be computed. Furthermore, behavior on singular points has to be taken into consideration.

The *transpose* technique utilizes a scaled transpose of the Jacobian $J^T K$ in order to solve the inverse problem. The solution $d\mathbf{q} = J^T \cdot K \cdot dM$ can be interpreted as a string with stiffness K pulling on the end effector [SS96]. This approach is computationally inexpensive, however, has often poor convergence characteristics or even no convergence at all if $K \cdot dM \in \ker(J^T)$. A discussion of the gain matrix K and overall properties of this approach can be found in [Wel93]. Optimizations of this method are described in [SH98].

The *pseudo inverse* technique uses $J^+ = J^T(JJ^T)^{-1}$ as the inverse. For a given change in the end effector dM, the joint rate $d\mathbf{q} = J^+ \cdot dM$ is a minimum norm solution, i. e., $||d\mathbf{q}|| = \sqrt{\sum_i d\theta_i^{x\,2} + d\theta_i^{y\,2} + d\theta_i^{z\,2}}$ is minimized. The general solution for $d\mathbf{q}$ is given as

$$d\mathbf{q} = J^+ \cdot dM + (I - J^+ J) \cdot \mathbf{y} .$$

The homogeneous part of this equation, the null space, does not affect the achievement of the goal and only contributes to the configuration. It can be, therefore, used to incorporate additional constraints and to optimize $d\mathbf{q}$ in terms of realism of the final posture and motion. Several authors have used the projection of an arbitrary vector $\mathbf{y}$ on the null space to improve the inverse kinematic computation. A general approach for inclusion of additional prioritized tasks is described in [NHY87]. Aydin and Nakajima use the secondary task $\mathbf{y}$ to describe balance of the virtual humanoid [AN99]. Yamane and Nakamura propose a solution for multiple end effectors with concurrent kinematic chains [YN03]. Their pin and drag algorithm handles multiple movable end effectors and constraints defined through fixed end effectors. The

[5] unless for well-defined kinematic chains with 6 DOFs

null space is used in order to integrate the whole humanoid body configuration in the solution. Baerlocher *et al.* present a technique for control of human posture in [BB98a]. These ideas are extended and incorporated into a system capable of handling multiple tasks that are encoded in the null space [BB04].

The *damped pseudo inverse* approach is a common technique for solving ill-conditioned equations in the area of kinematic simulation. The definition of the pseudo inverse is combined with a damping factor. The rotational configuration is computed with $J^* = J^T(JJ^T + \lambda I)^{-1}$, *cf.* [NH86, Mac90]. This approach omits singularities, however, the convergence is not as good as with simple pseudo inverse methods. The larger the damping factor λ the better the behavior near singular points; with λ close to zero the damped approach delivers solutions similar to the ones obtained with J^+.

Matrix Inversion. In order to solve the inverse problem with the (damped) pseudo inverse approach either $J^+ = J^T(JJ^T)^{-1}$ or $J^* = J^T(JJ^T + \lambda I)^{-1}$ have to be computed, respectively.

The straightforward approach is to invert $JJ^T \in \mathbb{R}^{6\times 6}$ or $JJ^T + \lambda I \in \mathbb{R}^{6\times 6}$. Apparently, an *analytical solution* for matrices of this size can be derived. However, as discussed in [MM04b] the analytical inversion of a 6×6-matrix requires more operations than an algorithmic decomposition of the matrix. Meredith and Maddock propose to use the *LU decomposition* and state that this method requires less then 10% of floating point operations in comparison to the analytical inversion. A disadvantage of both techniques is, however, that they do not allow for a direct control of singularities.

A widely used approach is the *singular value decomposition* (SVD), e. g., discussed in [Mac90]. A matrix $A \in \mathbb{R}^{m\times n}$ is decomposed such that $A = UWV^T$ where $U \in \mathbb{R}^{m\times n}$ and $V \in \mathbb{R}^{n\times n}$ are column-orthogonal matrices and $W = \operatorname{diag}[w_i] \in \mathbb{R}^{n\times n}$ is a diagonal matrix with positive or zero elements w_i that are the singular values of A. Given the SVD of the Jacobian $J = UWV^T$ the pseudo inverse is computed as $J^+ = V \cdot \operatorname{diag}\left[w_i^+\right] \cdot U^T$ with

$$w_i^+ = \begin{cases} \frac{1}{w_i}, & w_i > 0\,, \\ 0, & \text{else}\,. \end{cases}$$

The damped pseudo inverse is given as $J^* = V \cdot \operatorname{diag}[w_i^*] \cdot U^T$ with

$$w_i^* = \begin{cases} \frac{w_i}{w_i^2+\lambda^2}, & w_i > 0\,, \\ 0, & \text{else}\,. \end{cases}$$

This representation shows clearly the previously mentioned fact that for $\lambda \to 0$ the results of the damped pseudo inverse are similar to those obtained by the pseudo inverse technique. Furthermore, this approach allows for a distinct control of singularities, i. e., in the case that w_i is close to or equal 0.

Approaches based on the Jacobian allow for computations of arbitrary kinematic configurations and support the formulation of additional prioritized tasks. However, the computation often suffers from singularities and solutions obtained are limited for ill-conditioned formulations of the problem, *cf.* discussion in [Mac90]. Other techniques derive the solution from a geometric analysis or reformulate the problem as an optimization task. They are discussed in the next paragraphs.

Geometric Approach. Geometric approaches derive the solution for the problem of inverse kinematics from geometric observations. Wang and Chen use the cyclic coordinate descend (CCD) method (*cf.* [Lue03]) with recursive variations of joint rotations in order to find the best approximation towards the desired end effector transformation M_d [WC91]. Starting at the end effector, each joint in the kinematic chain is being regarded in a recursive order. The current orientation of the ith joint q_i is optimized in terms of two criteria, a positional and a rotational error. The positional error is defined as the alignment of the vectors $\mathbf{v}_i(q_i)$ and $\mathbf{v}_i^d$ that describe the current direction to the end effector from the joint's origin and the direction to the desired end effector position also with respect to the joint's origin. The orientational error is described by means of alignment of the rotational axes of the current end effector and the desired orientation. This approach is discussed and detailed in [Wel93, Lan98a].

The CCD-based method omits the time intensive computation of the Jacobian and its inversion and is, therefore, well suitable for fast results. However, it is limited to the end effector transformation and does not allow for the inclusion of additional constraints.

Formulation of Kinematics as an Optimization Problem. The idea of regarding inverse kinematics as an optimization problem is mentioned in [BPW93] and detailed in [ZB94, Zha96]. It has been proposed to use nonlinear programming in order to compute a solution for the inverse kinematics problem that satisfies a variety of spatial constraints. The problem is described as follows. Given a goal function $\mathrm{G}(\mathbf{q})$, find a $\mathbf{q}^*$ such that

$$G(\mathbf{q}^*) = \min\{G(\mathbf{q}) \mid \begin{array}{ll} \mathbf{a}_i^T \cdot \mathbf{q} \leq b_i, & i = 1, \ldots, l \\ \mathbf{a}_j^T \cdot \mathbf{q} = b_j, & j = l+1, \ldots, k\} \end{array}$$

with $[\mathbf{a}_i, b_i]$ representing the upper and lower limits for each joint, and $[\mathbf{a}_j, b_j]$ polygonal constraints for complex joints, like the shoulder. The goal function is defined as a weighted sum of subgoals. The basic goals are expressed as energy functions [WFB87]. Badler and Zhao derive representations and derivatives for several spatial goals, e. g., position, orientation, and aiming goals, as well as line, plane, and half-space constraints. The optimization problem is solved with a nonlinear solver[6].

[6] [BPW93] and [ZB94] use a combination of the BFGS and Rosen's projection method. [Zha96] suggests the use of the L-BFGS-B algorithm [ZBLN97].

Optimization techniques are a powerful approach towards inverse kinematics. They allow the definition of various constraints for end effector position and orientation, as well as a control of joint limits. The drawback of this approach is, however, that external solvers have to be utilized that have no knowledge about the particular problem. This lack of specificity makes the optimization process costly in terms of processing time.

1.1.5.2 Motion Synthesis

Motion synthesis is dealing with motion sequences, their synthesis and adaptation for animation of virtual humanoids.

At best, existing motion data can be directly applied to a suitable skeleton through forward kinematics. Motion data are mostly represented as a definition of a kinematic structure and a sequence of keyframes that consist of joint parameters and timing information, e. g., [SB85]. These parameters are interpolated in a meaningful way, e. g., basic slerp interpolation between two quaternions [DKL98] or more advanced interpolations with quaternion curves based on multiple control points [KKS95]. A quite often used technique for modification of particular keyframes within a sequence is *motion warping* [WP95]. It allows for the creation of smooth animations. Other editing methods are, e. g., motion retargeting [Gle98], editing of the trajectory (path) of a motion [Gle01a], approaches based on signal processing [BW95], physical constraints [PW99, TK05, LP02, ALP04, EAPL06], and style-based adaptations [PB02, GMHP04, HPP05]. An extensive survey of motion editing methods is published in [Gle01b].

Problems during motion animation arise if a stored motion sequence has not been recorded for the skeleton used, or if a desired motion is not contained in the motion data base and has to be synthesized from available sequences. In order to handle these problems several approaches have been discussed in the areas of adjustment and concatenation of motion, which are presented in subsequent paragraphs.

Adaptation of Motion Data to Kinematic Structures. The problem of motion adaptation arises if motion data have been (pre-)recorded or generated for a particular figure and have to be used with a different target. The main approach is to use the original motion data and to enrich them with additional constraints. These constraints are used to modify the motion. The modification is done on the basis of the previously described motion editing techniques. In addition the methods take a modified skeleton configuration into account.

Gleicher proposes a method that allows to retarget motion between kinematic structures with an identical topology but different offsets between joints (length of segments) [Gle98]. The retargeting problem is formulated as a large optimization problem. Gleicher uses a set of geometric constraints that can affect the whole motion sequence and not only single frames. These spacetime constraints, firstly introduced in [WK88], allow for creation of motions without high-frequency errors

that are typical for frame-based adaptations. This method also depicts the problem that an adaptation should preserve some characteristics in the new motion. These characteristics could address the style of a motion, particular spatial goals, like a human holding an object or remaining on the ground and not levitating in space, or biomechanical requirements. Gleicher addresses several geometric constraints [Gle98]. A quite important category of such constraints are given by the placement and motion of feet. Restrictions of motion of planted feet (footskating) are handled, e. g., in [KSG02]. Bindiganavale and Badler address the problem by analyzing the second derivative of motion trajectories to identify important keyframes and by applying inverse kinematics to enforce resulting constraints [BB98b]. Popović and Witkin present a method that allows to retarget motion between articulated structures with different topologies. Their method is based on spacetime constraints for dynamics and a simplified intermediate abstract character representation [PW99]. Lee and Shin use a multilevel frame-based approach for motion retargeting [LS99]. They apply inverse kinematics on end effector trajectories derived from the original motion in order to describe constraints in each frame. The new motion is computed hierarchically and approximated by multilevel B-splines in the time domain. The authors emphasize that their B-spline approach allows for a decomposition of the overall spacetime problem into manageable subproblems. A similar per frame approach is described in [TK05]. Similar to Popović's approach Monzani and colleagues use an intermediate skeleton for motion retargeting [MBBT00]. They extend the original idea in that motion for characters with kinematic structures with different joint offsets and topology can be processed. A method based on inverse kinematics and the tracking of end effectors is presented in [CK00]. They apply a closed-loop inverse kinematics scheme, e. g., [CCSS91], and use the end effector constraints as the primary task and the original motion as the secondary task. Even though computations are performed for each frame no ruptures in motion occur because of the continuity of constraints. In a more recent approach the ideas of inverse kinematics for end effectors and style of the original motion is combined in a system that dynamically determines the importance of end effector data and posture information [SLSG01]. These two data sources may lead to conflicts especially for spatially differing skeletal configurations. Shin *et al.* use a metric based on the object distance. They use a Kalman filter for noise reduction and use a combined approach for inverse kinematics similar to the one previously described in [LS99]. The system is used for on-line processing of motion capture data. Gao and colleagues present a technique for retargeting a set of motions to one skeletal configuration [GMCW05]. In a first step they adapt the motion to the new skeleton and perform a footskate clean-up as described in [KSG02] (*cf.* discussion above). The presented method is basically equivalent to the previously described technique by Gleicher [Gle98] but avoids a formulation of geometrical constraints.

The problem of motion retargeting is handled in a wide range of papers. Most approaches propose the use of constraints and propose to minimize incongruities in motion through additional adaptations. The reduction to spatial constraints ensures

certain desired end effector transformations, e. g., foot placement on the floor, however, it does not address the problem of differences—topology or joint offsets—of the underlying kinematic structure, in general. In most approaches a correspondency and alikeness of the two skeletons is presumed. The aspect of skeleton adaptation is handled by Monzani *et al.* and Gao *et al.* to a certain extent, but algorithms do not handle the twisting of bones, which is important for several motion data formats. Furthermore, most methods require a non-interactive processing of the motion. This aspect is crucial for the use of interactive virtual humanoids in virtual environments. Once the original motion is retargeted for a given skeleton the question arises how additional motion can be synthesized in order to reach desired goals, postures or movements that are not included in the original data set. These methods are discussed in the next paragraph.

Concatenation Techniques. In the majority of cases the pre-recorded and retargeted motion sequences will not be capable of representing the desired final motion. For navigation tasks, for example, generating all possible motions that would cover the configuration space would result in a large amount of motion data difficult to manage and handle. Therefore, the idea has been expressed and researched to combine parts of motion sequences that yield the desired animation sequence.

The quality of a concatenated motion is dependent on the transition technique chosen to blend two motion snippets. Without loss of generality, it is assumed that the parts considered are motions for one and the same skeleton or retargeted otherwise. Given two keyframes a transition could be simply generated by transforming the poses into each other. However, this leads to unnatural results if both poses differ much. Several authors propose ways to find and to generate optimal transitions between two sequences. Rose *et al.* present a method for generation of transitions between motion sequences that are dynamically plausible [RGBC96]. Their method uses spacetime constraints and optimization-based inverse kinematics. Lee *et al.* define transitions between frames by means of a multi-level model of the motion data [LCR+02]. On the basic level they define distance metrics for frame transitions and use them in a matrix, which represents a Markov process. On a second level similar poses (frames) are grouped into clusters. Each frame is provided with a local tree describing the connectivity of reachable clusters. These data are used to compute the most probable motion sequence. The cost metric proposed by Lee *et al.* is further analyzed by Wang and Bodenheimer [WB03]. They extend the original approach in that optimal weights are computed instead of estimations, as proposed by Lee *et al.*. This metric is evaluated in a set of studies and used to find the most suitable keyframes for blending in a set of motion sequences. Transition generation between two motion sequences is also discussed by Kovar and Gleicher. They propose a method for flexible automatic motion blending with registration curves [KG03]. These methods show ways how to blend two motions and how to find the best suitable keyframes. To achieve a specific goal by concatenating mo-

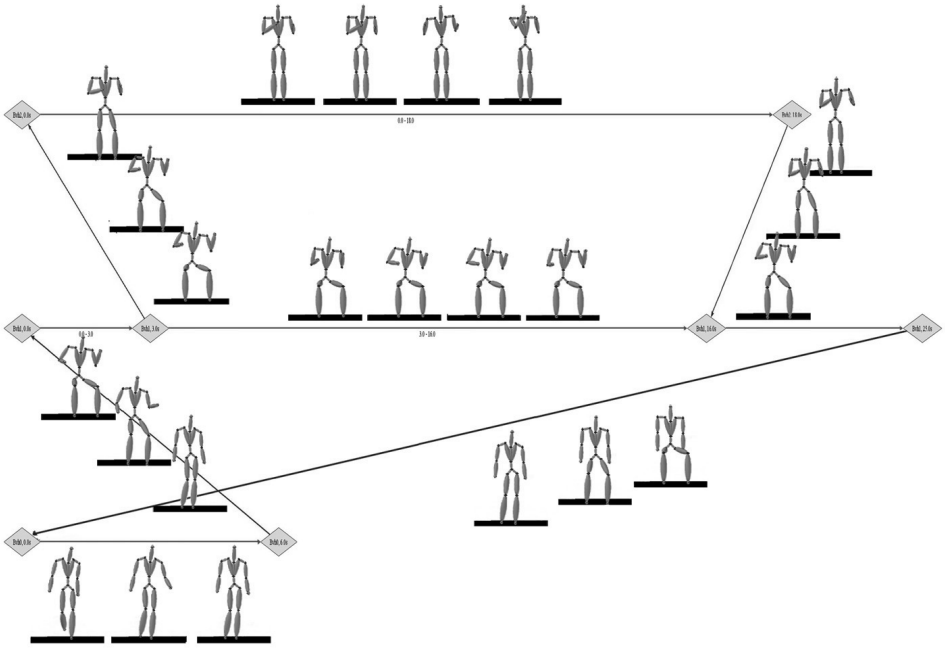


Figure 1.3. Motion graph with three original motion sequences (horizontal edges) connected with four transitions.

tion sequences the original data have to be organized in some structure allowing for specific inquiries.

Wiley and Hahn discuss the synthesis of motion from data bases [WH97]. The authors describe their approach as an extension to simply selecting and animating a motion. Another structuralized approach is described as move trees, e. g., in [MBC01]. Move trees are manually created graphs that are focused on simplified actions, like "move forward" or "turn left". They have been mainly used in computer games. More elaborated approaches propose the use of graphs for synthesis of motion [AF02]. Kovar *et al.* introduce the term motion graph [KGP02]. The idea is further extended in [AFO03]. The nodes of such graphs represent poses or keyframes, and edges particular motion sequences. The graph is assembled in that nodes are connected that include similar poses (*cf.* Fig. 1.3). An animation is equivalent to a path in the graph. Graph-based approaches describe motion synthesis as a graph search problem or as a computation of the best suitable path. However, they often do not allow for a detailed control of style and other features. Park *et al.* have used a graph approach with selected transitions [PSS02]. The interpolation between two adapted original motions is parameterized allowing for a higher control of the resulting style of the motion. Gleicher and colleagues present another view on the data base-driven synthesis [GSKJ03]. A user has to identify often occurring poses in the original set of motions. These poses are used to create a motion graph out of the original motions. In this preprocessing step the motion data is modified

in that all on-line computations consist of a concatenation of motion clips, only. This allows for a fast processing. The application of the motion graph is similar to the previously mentioned methods. The already cited work by Pullen and Bregler [PB02] uses several keyframes and derives intermediate poses from a motion data base. The method allows for a control of the style of the resulting motion. Brand and Hertzmann use hidden Markov models to learn motion patterns from a set of motions [BH00]. Li and colleagues also use a statistical model for the analysis of pre-recorded motions [LWS02]. The analysis yields a description, called motion texture. The algorithm determines connectivity between motion textures based on a maximum likelihood estimation. Instead of using one single motion graph for the entire environment Lee and colleagues propose to discretize the simulated environment into regions with specific motion possibilities [LCL06].

The main advantage of methods for motion synthesis is their capability of generating streams of motion of arbitrary length. If a motion graph has sufficient connectivity, it is capable of generating fluent animation streams for a variety of purposes. However, the performance, i. e., the range of motions that can be created with a motion graph, is still an open issue. Motion graph methods evaluate the quality of the resulting motion with local metrics, i. e., they guarantee a certain amount of quality in the transition. The previously mentioned method by Arikan and colleagues [AFO03] allows for some definition of motion styles by means of annotations, but their approach enforces high costs in preprocessing, i. e., each part of a motion sequence has to be annotated in advance. In [RP04], motion graphs are discussed in terms of global metrics and the authors present a method for unrolling of the motion graph to evaluate navigation of virtual humanoids. However, an analysis of construction techniques in terms of capabilities of the resulting motion graph has not been presented so far.

1.1.5.3 Motion Tracking

This section discusses methods for acquisition of motion data and focuses on on-line capable techniques that are of interest in virtual environments. Motion tracking is dealing with the mapping of the kinematics of an actor (or user) on an articulated structure, i. e., the creation of motion data described in the previous section. Either this process is done offline, or directly applied on the kinematic simulation of a virtual humanoid. The former is generally referred to as motion capture or performance-driven animation, while the latter is set in the domain of direct interaction.

Capture. Methods for offline acquisition of motion data are referred to as motion capture techniques. A basic introduction to this topic in the area of education is presented in [Ger04]. There are two general approaches: the use of image-based methods and the use of markers.

Image-based methods derive the motion from a series of images that are analyzed in order to determine a human contour and the current pose, e. g., recently published work in [RSH+05, SMH05]. These methods are highly flexible, because they do not require any marker set-up. However, further assumptions have to be made, like skeletal configuration and information about environmental setting, in order to be able to recognize the human contour.

Approaches utilizing markers require that an actor is connected to several sensors that are placed on his body. In an offline step, the data is processed so that they determines a pose. Several commercial motion capture devices are available, e. g., by Vicon[7], Elite [FP85], or MotionAnalysis[8], to name some. These devices record trajectories of markers, and allow for some control of occlusion and identification problems. Skeleton-based motion data is computed in the final step.

The whole motion capture procedure is described in [BRRP97]. An approach that increases the robustness of motion capture by already considering a skeleton during trajectory computation is reported in [HFP+00]. A similar idea is presented in [CRPF03, CPF05]. Here, a kinematic model is applied in order to reduce artifacts from displacements of markers on the skin. A method for skeleton fitting is described in [SPB+98]. Zordan and Van Der Horst describe a system that uses computation of dynamics in order to apply sensor data to a skeleton [ZV03]. Sensors are viewed as external forces pulling on a geometrical hull of the skeleton. This general approach allows for a definition of additional constraints, such as, positioning of feet and internal forces describing an initial pose. The system is capable of producing 2–3 frames per second and, therefore, can not be viewed as interactive. In a recent work Park and Hodgins present a method for capturing deformations of skin [PH06]. They derive motion data from a large set of markers and combine this information with a surface model for skinning, i. e., the visual representation. Liu *et al.* study the redundancy of tracking markers [LZWM06]. They analyze motion capture data to determine sets of significant marker. These are used to compute the final pose of the tracked character.

Online Motion Tracking. Online tracking methods are closely related to motion capture approaches described in the last paragraph. Tracking techniques use a number of sensors to deliver positional and eventually orientational data, and derive a skeletal configuration based on the sensor data. The main difference between tracking approaches and motion capture techniques is that tracking is often limited by the amount of sensors used and the fact that tracking data have to be computed in real-time while motion capture data is processed after recording. Another issue of tracking is usability. The use of a sensor should be straightforward without time-consuming calibration procedures. In addition, the time to configure the sensors on the user's body, the so-called don-doff time [CH05], must not exceed a certain duration, or the whole procedure is perceived as awkward.

[7] http://www.vicon.com, last visit May, 2007.
[8] http://www.motionanalysis.com, last visit May, 2007.

Some approaches focus on selected body parts, like limbs or the head, in order to achieve interactive frame rates. An example is the work of Jorissen and colleagues, who develop an approach to arm tracking within their collaborative, distributed virtual environment [JWL05]. In an early work Badler *et al.* use four sensors for the tracking of the head, the upper body, and the arms [BHG93]. More general methods try to achieve interactive processing by omitting positional sensor data and inverse kinematics methods. Molet *et al.* propose such a method focusing on orientational data from electro-magnetic trackers [MBT99]. The previously cited work by Shin *et al.* also incorporates an online tracking method in their computer puppetry approach in order to determine spatial constraints generated by actors [SLSG01]. They describe spacetime constraints and inverse kinematics to create motion with the desired features. Unfortunately, the tracking procedure is not detailed in the paper. These methods require a predefined mapping of the tracking sensors to a kinematic structure and, furthermore, a calibration phase, where the tracking system is initialized and linked to the virtual skeleton.

This process is avoided by self-calibrating methods that automatically derive the underlying kinematic structure. A method for automatic assignment of trackers to segments and related joints of an automatically determined skeleton is described in [OBBH00]. This approach is extended to a self-calibrating system for virtual environments by Hornung *et al.* [HSK05]. The method uses one tracker per segment without prior knowledge of the skeletal set-up and the initial configuration of the joints. After an initialization phase tracker data is used for automatic computation of the skeleton configuration and subsequent determination of joint positions. On the one hand, this method allows for the use of arbitrary skeletons, like limbs, other parts of the body, or even non-human configurations. On the other hand, this indetermination results in quite inaccurate and fluctuating offsets between consecutive joints.

The detection and handling of discontinuities caused, for example, by sensor drop-outs or biased data, is a difficult and often not handled aspect of online motion tracking methods. As online systems can not analyze the future motion data they have to rely on past data and heuristics to detect and to correct possibly erroneous data. Tracking data can be analyzed with predictors and filters, like the extended Kalman filter discussed, e. g., in [vRvLM05]. Such filters are often already implemented in the tracking system itself. In order to improve the robustness of the final tracking data it has been proposed to combine physics simulation and motion data. Hodgins and Popović discuss the application of simulation of physics on motion capture data [HP00]. An interactive tracking method that includes simulation of dynamics is presented in [ZH02].

1.1.6 Musculature Simulation

It has always been one goal in research on virtual humanoids to simulate internal structures. One focus has been to simulate physiological procedures, the other to

create life-like appearance of organs and body parts. On the one hand, immediate applications can be found in various medical areas, e. g., VR-based simulators and teaching tools. On the other hand, results from simulation of internal structures are used for improvements of realism in visualization, and kinematics. Musculature is a prominent example for simulation of internal structures. Musculature has a high topological, functional, and morphological complexity and is an essential component in all body parts. Because of the close link to the musculoskeletal system, it has been proposed to use results from the simulation for improvements of kinematics, e. g., the muscle-based inverse kinematics approach described in [KSK00, KSK01]. In addition, musculature can be used for improvements of visualization (*cf.* section 1.1.3.3). In general, simulation of musculature focuses on two aspects: the forces of the muscular system and musculature deformation. The following sections detail current approaches.

1.1.6.1 Simulation of Forces

The biomechanical simulation of musculature is mainly divided into two categories. In forward dynamics trajectories are computed for given forces, whereas inverse dynamics is used to compute forces for given movements. For this purpose muscles are usually represented by lines of action [BCW+82, Hat76, Har78]. These lines are either direct lines (action lines) [SA89] that connect the origin and the insertion of the particular muscle, or curved lines that follow the centroid of the muscle [JD75]. Polylines are often used as discretized centroid-lines [Rai92]. The simplest way to approximate muscle forces is to use a linear spring-mass model that solely relies on muscle-length deflection and an empirical estimated stiffness coefficient. A more complex model is created by Hill [Hil38, Win90]. In this approach a muscle (or its action line) is represented by a combination of parallel and serial springs and a contractile element. An extension of the Hill model is presented by Zajac [Zaj89] that additionally takes the alignment of muscular fibers into account. In a recent approach, Lee and Terzopoulos, presented a system for the muscular control of the neck [LT06]. They have used a neuromuscular model and solved the motor control problem for 72 affected muscles.

These approaches allow for computation of muscle forces and its spatial length. The computation of the overall spatial deformation of the muscle geometry is often omitted for the sake of time efficiency. Approaches that solve this task are described in the next section.

1.1.6.2 Visualization and Deformation

There are several approaches to deformation and representation of musculature. The first methods were developed in the area of biomechanics [DL95]. In this field it is common practice to render simplified action lines that are used to simulate the kinematics and dynamics of muscles (*cf.* Fig. 1.4). More advanced approaches use basic geometric shapes to represent musculature like volume-preserving ellipsoids

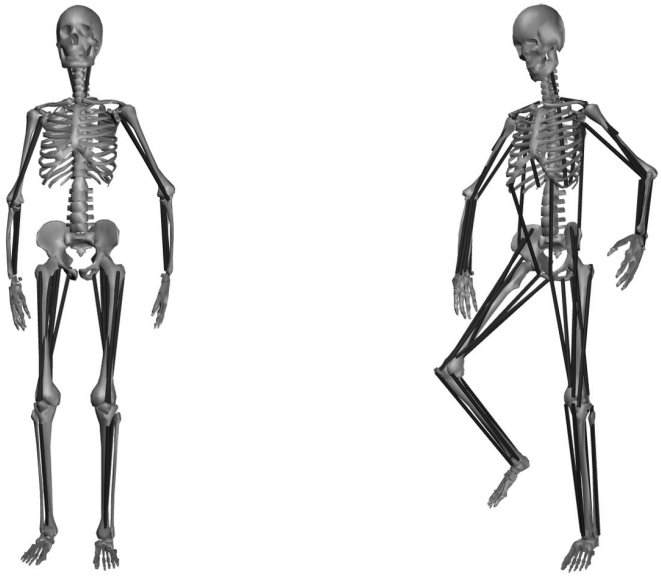


Figure 1.4. Musculature defined by action lines.

[SPCM97], or deformable cylinders [WG97]. In both cases the length of the muscle is used to scale the representing geometry in a volume-preserving way. Volume-preservation is often discussed as a factor of realistic visualization of muscle geometries. Even though muscles do not strictly preserve their volumes during relaxation and contraction the volume remains approximately the same because they do not loose or gain mass. Another technique utilizes merged, implicit spheres, so-called metaballs [Bli82, NHK+85, WMW86]. For visualization of musculature Thalmann *et al.* provide a heuristic for modification of parametric sets dependent on the angle of the parent joint and thereby deforming the metaballs [TSC96]. NURBS that are quite popular in computer-aided design are used by Dong *et al.* [DCKY02] in order to construct muscle-geometry with contour-lines extracted from the visible human project [SASW96]. In a similar approach, Ng-Thow-Hing and Fiume introduce B-spline solids [NF02]. They propose a new tri-variate function that supports control points inside the enclosed volume [NF97] and state that this makes adjustments easier. Many interesting approaches involve detailed representations of muscle anatomy, e. g., [LTW95, NT98]. Looking into the field of volume-rendering there are some solutions adding dynamics through classical finite element (FE) methods [CZ92, ZCK98]. In a recent approach Teran *et al.* [TSIF05] propose an extended and robust FE technique coupled with a dynamic free-form deformation embedding scheme that reduces computational costs. Hong and colleagues present a spring-mass system for deformation of musculature [HJCW06]. The authors claim that their system achieves results similar to those of FE methods.

The techniques described so far have a coarsely ascending degree of realism, as well as computational costs: they either utilize highly detailed and realistically deforming geometries to visualize musculature and accept non-interactive frame rates;

or, simplified low-resolution geometries like geometric primitives are used in order to achieve fast results.

1.1.7 Behavior and Cognition

Behavior and cognitive simulation is a broad research area and is addressed by several toolkits mentioned in section 1.1.1. Even though simulation on this level is not the focus of this thesis basic approaches are briefly sketched. The humanoid entity is often regarded as part of a virtual world. The humanoid must be able to perceive the world and to accomplish specific goals. While behavior simulation mainly focuses on the appropriate reactions to environmental stimuli, cognition requires an analysis of the surrounding events and goals, planning, and the drawing of conclusions that are, finally, transformed into quasi-intelligent actions in the virtual world. A further description of cognitive modeling is given by Funge *et al.* [FTT99]. The authors state that cognitive models "govern what a character knows, how that knowledge is acquired, and how it can be used to plan actions". Another definition emphasizes goal-oriented "directability" at the motivational and task level of an autonomous virtual humanoid [BG95]. This brief record gives general ideas concerning simulation of behavior and cognition. It should be emphasized that behavior and cognition are essential parts in the simulation of virtual humanoids and that approaches to integrated systems must allow for seamless integration of such methods.

1.2 Discussion of Methods

So far, systems and toolkits, as well particular simulation approaches for virtual humanoids have been presented. Current techniques for visualization, speech processing, animation of motion, and simulation of musculature have been described in great detail. A short insight into the simulation of behavior and cognition has been presented, as well.

The majority of virtual humanoid toolkits as presented in section 1.1.1 use an articulated structure as their basis for kinematic simulation and animation. Visualization focuses on skin and cloth rendering and in some cases it is restricted to morph targets for facial animation. In addition, extensive efforts have been put into researching autonomous and interactive behavior in virtual humanoids with the goal to create virtual learning and tutoring scenarios. When comparing the functionality of the systems, one needs to take into account the fact that many of the existing tools are highly optimized for particular applications. Moreover, as some toolkits are commercial products the source code is not freely available for adaptations. The specialization implicates that adaptations to varying requirements and tasks are difficult. As an example, medical applications require a flexible data structure capable of holding various data for physiological simulations of internal structures. Some aforementioned toolkits contain a soft tissue data structure, however, the inclusion of

additional organs and systems are subject to limitations and restrictions. It is, therefore, not possible to develop comprehensive simulations with these approaches. In addition, several simulation approaches require time intensive computations. When included in the aforementioned toolkits—if possible at all—they lead to a significant slow down of representation and interaction. The reasons are mainly missing methods for distribution. Such network capabilities are of importance also for distributed virtual environments and, therefore, essential for application of interactive avatars in communication scenarios. Lastly, an application of virtual humanoids in virtual environments requires elaborated, task-specific kinematic interaction and tracking. This means that a low level access to the kinematic structure, in most cases the articulated joint structure, has to be provided. Specialized toolkits often do not offer this capability. To sum up, the systems presented in section 1.1.1 allow for highly developed solutions for certain tasks that require virtual humanoids. However, an integrated solution with flexible and comprehensive data structures, concepts for integration of complex simulation approaches that maintain interactive frame rates, and possibilities for specific kinematic interaction and tracking algorithms does not exist yet.

Beside the required data structures and simulation interfaces provided by a toolkit, several methods for simulation of virtual humanoids presented manifest considerable potential for improvements. In particular, there are not yet satisfactory solutions for synchronization of immersive, multimodal representation, synthesis of motion, whole-body tracking, and realistic deformation of musculature.

The methods for visualization and for speech synthesis as presented in sections 1.1.3 and 1.1.4, respectively, have shown a broad range of approaches for the representation of virtual humanoids. Both modalities are integrated in TTAVS systems, however, current systems do not support immersive representation well. In particular, a system for synchronized representation of audio-visual speech with methods of binaural synthesis and stereoscopic visualization for immersive virtual environments has not been presented, so far.

Motion synthesis is essential for animation of kinematics for virtual humanoids. The approaches presented in section 1.1.5.2 handle the complex task of utilizing a set of pre-recorded motion data to create new animations. However, they are limited in several aspects. Either they are capable of utilizing and adapting a restricted amount of input sequences and generate a limited set of resulting motions, or they use motion graphs allowing for a wide range of results, in theory. Yet the quality of motion and connectivity is handled on a basic level, only. An analysis of parameters for construction of motion graphs in terms of navigation tasks and quality of resulting motion is still missing. Such data would allow for a fast construction of motion structures with optimal navigation characteristics.

Online motion tracking as described in section 1.1.5.3 represent an essential input technique for interaction with virtual humanoids. Classical capture methods are unwieldy for application in interactive virtual environments because of their latency issues and the fact that data have to be preprocessed before being applied

on an articulated structure. With regard to the task of whole-body interaction with virtual humanoids direct motion tracking methods have to be used. However, they are mostly focused on tracking of trajectories used as end effectors for inverse kinematics for parts, or limbs of the humanoid. So far, a limited number of integrated systems for virtual environments has been proposed, e. g., in the basic work by Molet *et al.* [MBT99], or Hornung and colleagues [HSK05]. Hornung's approach has the advantage to automatically determine the underlying articulated structure. This is an interesting feature and leads to a fast application in, e. g., animals; however, as humans are the most likely users of such tracking systems and given the fact that the human skeleton is well known it would be promising to factor the skeletal configuration into the determination of the tracked pose. Moreover, the quality of the original tracking data is not handled in most methods. Unfortunately, it is a common situation that tracking systems are interfered and the resulting data biased. Consequently, the robustness of online tracking systems has to be researched and improved.

Simulation of musculature has gained greater attention due to its application for kinematic simulation and visualization with improved degree of realism. Nevertheless, most approaches are not applicable to interactive virtual humanoids because they require time-consuming processing. Interactive approaches are often bound to simplified geometries or selected parts of the body, like the muscular system of the upper arm. A flexible, yet interactive deformation of the overall human musculature is still lacking.

In summary, the discussion of the prevailing state-of-the-art in toolkits and simulation methods for virtual humanoids motivates further research in the areas of general, extendible data structures and flexible data processing, as well as methods for synchronization of immersive, multimodal representation, strategies for synthesis of motion, robustness in online body tracking, and realistic deformation of musculature. These conceptual points are set forth in chapter 2. Detailed solutions and their evaluation are presented in chapter 3.

1.3 Presence in Virtual Environments

Virtual environments are useful working surroundings, for example, to explore large data sets, to handle non-existent or dangerous objects, or to accomplish collaborative and remote working tasks. These and further application scenarios illustrate the highly sophisticated interaction possibilities. Unfortunately, interaction within virtual environments can be a tedious task. In addition, inauspicious settings and scenarios may lead to more or less strong physiological reactions on behalf of the user, e. g., simulator sickness [KLBL93], disorientation, slowdown and inaccuracies of the working process, or simply a resulting negative attitude towards the virtual environment.

From a technical point of view this fact is basically caused by the still existing gap between reality and the computer-generated virtual reality. This gap may arise due to a number of factors such as the hardware realization of the virtual world (display or input devices), an erroneous and imperfect simulation of the objects within the virtual world, or non-transparent interaction possibilities with the virtual environment. We are still a long way from perfectly simulating reality in all its characteristics and the human brain is able to detect even the slightest inaccuracies in sensory data and content behavior. This process is not even solely dependent on the degree of realism. Virtual environments with highly detailed and realistic content, for example, can be perceived as even more unnatural. This effect is referred to as the "uncanny valley" phenomenon [Mor70, BGBC05]. Human-computer interaction and neuroscientific approaches closely link the overall problem to human factors, and cognitive and mental processing that are summed up under the term of presence.

It is, therefore, imperative to study methods that help to overcome the evident drawbacks of current virtual environments, the basic goal being to enhance virtual environments so that they resemble a transparent interface [Ish04, OC00] to a life-like virtual world. A promising way is to research interface techniques and design principles capable of enhancing the sensation of presence and the usability of virtual environments.

Subsequent sections introduce the concepts of presence, its interrelation with virtual humanoids used as avatars, and the influence of presence on usability.

1.3.1 Presence

The original idea of presence was anticipated in a novel by Robert Heinlein [Hei50][9] and reformulated as telepresence by Minsky [Min80]. Loomis discusses the close relation of presence and the phenomenon of distal attribution [Loo92]. Presence is viewed as a state of consciousness and part of the attribution of sensation to some distal stimulus, possibly generated by a mediated environment. Current definitions of presence, as given by [She92, Sch95, LD97, SW97, WS98], relate the topic to the experience of users in virtual environments. Sheridan defines the term "virtual presence" as a subjective experience or the illusion to be located in a virtual place [She92]. Witmer and Singer [WS98] use a similar definition:

> *[Presence is] the subjective experience of being in one place when one is physically situated in another.*

In [PPFW95] the authors formalize presence as

[9] Heinlein's fiction "Waldo" is about a scientist afflicted with *Myasthenia gravis*, who, therefore, lives in the weightlessness of the orbit. His main inventions are remote-controlled robots that fulfill tasks on the earth. The operator observes the action through a stereoscopic display transmitting the perception of the robot's eyes.

> *[...] an illusion of position and orientation: i. e., that presence has to do with a switch in the cues one uses to determine one's position and orientation, from using cues defined by the real environment to using cues defined by the virtual environment.*

Accordingly, most definitions of presence are simplified to the term "sense of being there" in a mediated environment. Steuer [Ste92] cites an unpublished manuscript by Reeves [Ree91] where the "sense of being there" is discussed in connection with consumption of TV media. Heeter [Hee92] also uses this term but does not explicitly connect it to presence. The work by Slater *et al.* [SUS94], IJsselsteijn and colleagues [IdRFA00], and the book [RDI03] edited by Riva *et al.* use the term more frequently and relate it explicitly to presence. Schuemie *et al.* mention the term "being there" in connection with measurement of presence [SvdSKvdM01]. All previous definitions focus on the fact that a user is situated in an environment interacting with it and perceiving it with his senses. Presence is understood as a subjective acceptance of the environment as the current location (*cf.* Fig. 1.5). Presence in a real environment is, therefore, defined, as well, and should be at its peak for healthy human beings. The concept becomes more exciting if the considered environment is generated through media and, therefore, variations of presence are to be expected. Another widely spread circumscription of presence is that of "suspension of disbelief", which is a formulation from the reverse perspective. It is equivalent to the previously mentioned definitions, however, it emphasizes the fact that inconsistencies in the perceived environment create a sensation of disbelief towards the environment. Presence is reciprocally identified with this sensation. Similarly, the disbelief should be low in reality and starts to vary in mediated environments.

Even though most authors agree that presence is a subjective sensation of a user experiencing media there is no unique, generally accepted definition of presence. The next section describes and identifies main factors influencing and causing presence that allow for some deeper understanding of the concept.

1.3.1.1 Types of Presence

Several types of presence are discussed in the literature. Heeter suggests three types of presence: subjective personal, social, and environmental presence [Hee92]. Personal presence is characterized by the feeling to be in the mediated environment. Social presence is given by interaction with other beings in the environment. Lastly, Heeter defines the environmental presence as a kind of behavior and reactivity of the virtual environment on the user itself. Schloerb distinguishes objective and subjective presence[10]. Objective presence is identified with task completion capabilities of the user. Subjective presence is characterized as the user's conviction that he is physically present in a given environment [Sch95]. Biocca discusses three types

[10]In fact, Schloerb's focus is on telepresence. However, conclusions can be drawn on presence, in general.

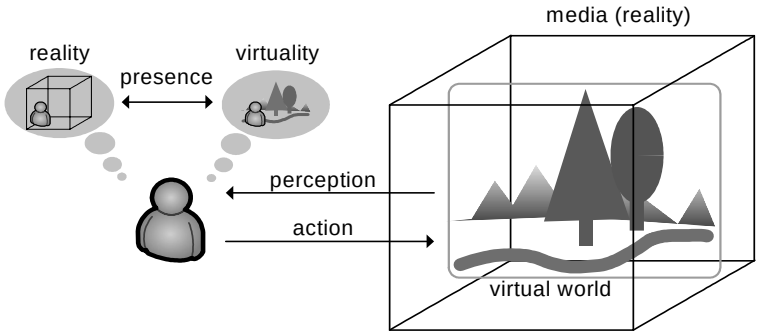


Figure 1.5. Presence defined as the subjective sensation of a user to be located in an environment (reality *vs.* virtuality), or as the suspension of disbelief against a mediated environment.

of presence: physical and social presence, as well as self-presence [Bio97]. Physical presence refers to the belief to be located in a particular place. Biocca points out that a user can feel physically present in one of three reference frames at one point in time: in the physical, the virtual, or the imagined environment. Social presence is given by communication with other beings in the mediated environment. Lastly, self-presence refers to the topic of embodiment and is defined as the effect of the virtual environment on the perception of ones body.

1.3.1.2 Factors Causing Presence

As discussed previously, presence is linked to interaction of a user with a mediated environment. It is, therefore, straightforward to classify the variables and factors influencing presence in external factors dependent on the media system used, and internal factors directly related to the user [SU93b]. Slater and Usoh define internal variables as the individual differences in how a user processes the information provided by the environment. External factors are related to the system generating the environment and are determined by the degree of realism of its content, i. e., how well the system is capable of replicating the reality in all its facets: the simulation and representation of content, and the interaction capabilities with the environment.

Internal, subjective factors are manifold. Sas and O'Hare discuss cognitive factors such as absorption, creative imagination, empathy, and willingness to experience presence [SO03]. The last factor has also been discussed under the term of involvement and attention towards the mediated environment. Sadowski and Stanney [SS02] review further factors: the influence of length of exposure to the environment, and social factors. Subjective factors depend also on the mental and physical capability of the user to perceive and process the presented stimuli. It is highly subjective, which particular sensory channel, i. e., visual, auditory, haptic, *etc.*, and which frame of reference, i. e., ego- or exocentric, the user prefers. A mismatch in

available and preferred sensory channels and frame of reference leads to a decrease of presence, as mentioned in [SU93a].

External factors are determined entirely by the hardware and software that drives the mediated environment. The degree of technical fidelity to create sensual stimuli resembling the real world is called immersion [SW97]. Even though Witmer and Singer used this term in a psychological, subjective sense [WS98], it is broadly accepted as an external factor relying on hardware and software. Several studies have concluded that presence is increasing in systems with a larger field of view, e. g., [LDA+02], or with a high pictorial realism, interactivity capabilities, and low delay of visual feedback [WBL+96]. In a recent study Sutcliffe and colleagues have compared three different virtual environments in terms of usability and presence [SGS05]. They found better usability and a higher sense of presence in large projection virtual environments. Plausible interaction and reactions of the simulated environment are influencing presence, as well [SUS94, SUS95, UAW+99]. The sensual stimuli generated by hardware are the most elementary information perceived that allows the user to create a mental representation of the mediated world capable of inducing the illusion of not being located in the virtual environment itself, but instead in the simulated virtual world.

Both, the internal and the external factors are essential for the experience of a mediated environment. A virtual environment can be technically realized with impressive fidelity, however, if the user is not poised to be involved, no large presence sensation can be expected. Similarly, an environment with very low technical quality is less likely to induce presence, as well. However, low hardware quality does not prohibit sense of presence, as presence can be invoked even in text-based media [TT97, Jac01, NB03].

Beside this general classification several authors propose characterizations based on factor analysis of questionnaire results. Schubert and colleagues have analyzed the factors of their questionnaire based on two studies with 246 and 296 participants, respectively, and concluded that three factors are dominant: spatial presence, involvement, and judgment of realness [SFR01]. While the most dominant factor spatial presence corresponds to the original definition of "sense of being there" the involvement factor is characterized by awareness and attentional processes. Judgment of realness is concerned with the degree of realism of the mediated environment. It is interesting to point out that spatial presence and realness have turned out to be two distinct factors in this study. This indicates that the influence of realism on presence is integrated in a distinct way presumably independent from the main factor spatial presence. In a recent work Witmer *et al.* have analyzed their questionnaire with a sample size of $N = 325$ [WJS05]. They determined four factors identified as involvement, sensor fidelity, adaptation/immersion, and interface quality. Beside the general involvement in the mediated experience, the involvement factor addresses the ability to control events of the mediated environment and judges the responses simulated by the environment. The focus of sensor fidelity is the quality of multimodal representation. Interface quality was used to evaluate whether control devices

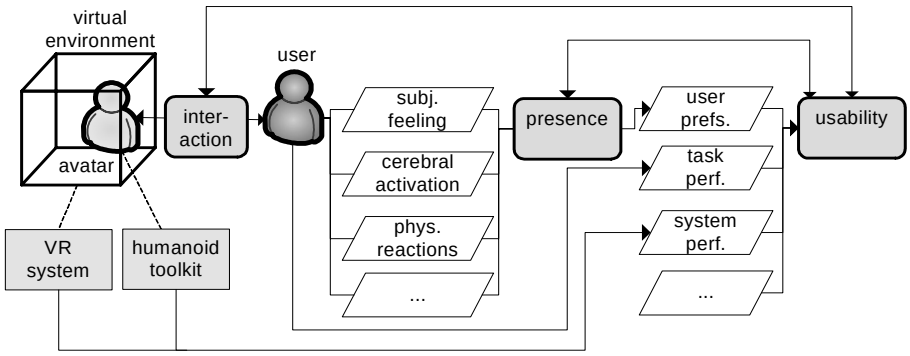


Figure 1.6. Interrelations between user-avatar interaction, presence, and usability.

or display devices are in line with task performance, and whether the participants felt able to concentrate on the aim of their work in the virtual world.

While these studies mainly reflect the capability of particular questionnaires to measure specific factors they also help to reveal essential components of the concept of presence, and express the distinction between external and internal factors. In contrast to the factors of involvement and spatial presence that reflect the internal subjective attitude towards the mediated environment, the factors judgment of realness, sensor fidelity, and interface quality represent external factors focused on software and hardware capabilities of the virtual environment.

1.3.2 Usability and Presence

While interaction or even imagined interaction is supposed to increase the sense of presence, e. g., [RS02], the extent to which a higher sensation of presence contributes to interaction performance [SS98, p. 163 *f*] or usability, in general, remains a topic of research. Bowman *et al.* mention that user preference metrics—and this term also includes presence (*cf.* Fig. 1.6)—generally contribute to overall usability [BKLP05]. They also emphasize that presence and user comfort can serve as important metrics for 3D user interaction techniques missing in traditional evaluations. Presence scholars discuss the topic in a controversial way (*cf.* the survey by Schuemie *et al.* [SvdSKvdM01]). Several studies have reported a correlation between subjective presence and tracking performance [EDM+97], search task performance [PPW97], and sensorimotor tasks [MAN97]. Psychomotor tasks are discussed ambivalently. While Witmer and Singer found evidence for a correlation with subjective presence [WS94], other studies by Singer *et al.* revealed none [SECP95] or found a positive impact of spatial knowledge on presence [SAMG97]. By contrast, Bailey and Witmer did not find any [BW94].

The previously discussed fact that the sensation of presence is closely associated with attention and involvement allows the assumption that some applications will not profit from an increased presence. In contrast, several environments may benefit from the increased attention and involvement: for instance, various platforms for

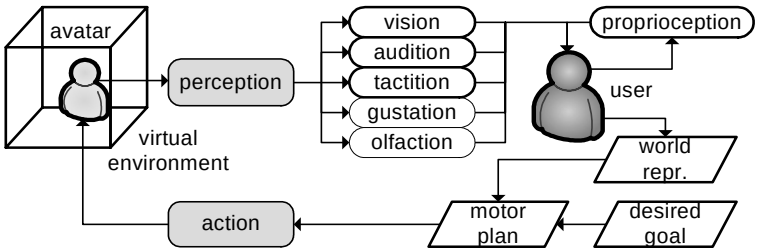


Figure 1.7. Perception and action. Sensory data perceived by exteroceptors are directly dependent on the representation in the virtual environment.

presentation (industrial large-platform design and presentation, design in automotive industry, architectural walk-throughs), surgical applications, neuropsychological assessment and treatment systems, entertainment platforms, as well as business and shopping systems. The goal of these applications is to draw the focus of attention on the content.

1.3.3 Avatars and Presence

The topic discussed in this section aims to answer the question whether avatars have the capability to enhance presence and general interaction metrics (*cf.* [BKLP05, Chapter 11]), in particular, task performance and user preference.

Avatars and self-embodiment are discussed by several researchers as one factor that increases presence, thought to be the result of a strong perceptual link between the mental representation of one's own body and the perceived virtual body [SU93b, Bio97]. In more detail, Biocca postulates three body representations: the natural (objective) body, the virtual body represented through an avatar, and the body schema, also referred to as body image [FC68], which is the user's mental concept of his body. This mental representation is not stable and can be altered in a mediated environment. In particular, the body schema is oscillating between the natural and the virtual body. Biocca emphasizes that avatars in immersive virtual environments can significantly bind to the body schema and have the potential to replace the natural body in the mental representation, resulting in an increased sensation of presence. This effect is observed even in comparatively primitive media systems.

In the majority of studies avatars and presence are mainly discussed in combination with HMD-based virtual environments or desktop-VR systems. Furthermore, mostly hand avatars are applied in order to study basic interaction tasks like manipulation, reach-to-grasp, and docking, as well as to evaluate sensory feedback interrelations. Although the hardware applied and the techniques for graphical representation have been somewhat limited in past studies, avatars are generally considered as a factor capable of strongly increasing presence.

In order to identify interesting factors influencing presence one has to analyze the user-avatar interaction and the related loop of perception and action, which is

sketched in Fig. 1.7. The human body can be described as an "array of sensors propelled through space to scan, rub, and grab the environment" [Bio97]. The user perceives the virtual or real environment and creates a mental world representation mainly provided by exteroceptors. Based on this information and the desired goal, a motor plan is generated and transformed into movements. This action is controlled on-line by sensory feedback and the motor plan is corrected if necessary. The consistency of data from different modalities affects the acceptance of the perceived environment. Mismatch between sensory information may even lead to the previously described symptoms similar or equal to simulator sickness. In the particular question of avatars and presence the acceptance of the virtual body—i. e., the identification of the body schema with the virtual body—depends likewise on the correlation of the perceived sensual data. In general, the user changes parameters of the avatar in a direct, synchronous way and perceives the virtual body representation through his senses. Beside the common human *external senses* of sight, hearing, touch, and in VR more theoretically regarded senses of taste and smell, an *internal sense*—the proprioception—has a large impact. Proprioception delivers data about the body posture and motion and is, therefore, essential in the projection of the body schema. The correlation between the external senses and proprioception is substantial for the acceptance of the virtual body.

While the believability of the avatar is judged based on stimuli registered by exteroceptors the effect of the motor plan and the consistency evaluation of proprioception are directly related on the simulation of kinematics in the virtual humanoid. As a result of these considerations two aspects of the simulation of avatars and their influence on the sensation of presence have to be further analyzed: the *degree of realism* of multimodal representation and the *kinematic* link between the user and the avatar. Related studies addressing these two factors are discussed in subsequent sections.

1.3.3.1 Realism of Representation

Several scholars have concluded that the degree of realism may increase presence. Mason *et al.* state that enriched graphical feedback about self-movement can be beneficial in tasks that require a high degree of precision and accuracy in the manipulation of virtual objects [MM04a]. In a recent study Durlach *et al.* report the positive influence of realism on presence, performance and even orientation [DFM05].

The degree of realism used for geometrical representation of avatars may substantially vary. In several studies simplified geometries are referred to as avatars, e. g., simple shoe geometries in [RS02]. Sometimes even geometrical primitives, like cross-hairs or circles, are used in order to represent position and motion [MM04a]. Simple representations that resemble a human body on a basic degree of realism are used, e. g., by [DFM05] for hand representation or by [SSUS00, VGSS04] as whole-body avatars. Slater and colleagues use avatars in order to research the social behavior of users in a collaborative virtual environment [SSUS00]. Interaction is

done on a basic level, mainly addressing positional control of the avatar. Opacity of the graphical representation during interaction is addressed in [BNB05]. Perani and colleagues compare the observation of movements of a real hand and those of virtual models in two different levels of realism [PFB+01]. In their functional imaging study they found overlapping cerebral activations. Because activations were not exactly the same in observation of the real hand and the two conditions employing virtual hand models, Perani *et al.* conclude that VR should only be used for psychophysical experiments with care. However, the results indicate interesting similarities in the two virtual conditions.

Another aspect of realism deals with similarity between a user and his avatar. This aspect is important, in particular, for teleconferencing and distributed virtual environments. Riva and Melis, for example, research variations of the body image of the users while perceiving their own avatar with varying body size in a virtual mirror [RM97]. Similarity is important in applications where recognition and acceptance of the interaction partner (or his virtual representation) is of major importance.

1.3.3.2 Kinematics and Proprioception

Kinematics is an essential factor in determining the influence of avatars on presence because it expresses the results of the motor plan and is involved in the integration of proprioceptive data. Kinematics in virtual environments is researched by several groups. Kuhlen and colleagues compare reach-to-grasp movements with real and virtual objects [KKS00]. This study has not involved virtual humanoids but has contributed to the clarification of the cerebral organization and control of movements. Several researchers have examined the particular influence of modalities and their variations on presence and performance [MM04a, DFM05]. Mason and MacKenzie suggest that for highly complex tasks, avatars with sophisticated kinematics could minimize task complexity. Kinematics plays a central role for the overall acceptance of the virtual body, i. e., the avatar, as a representation of oneself. Some functional imaging studies aim at kinematics of a virtual humanoid in order to evaluate cerebral effects. However, no particular functional imaging study has been reported that would research avatar-mediated motion and correlated aspects of synchronity. Several related studies deal with perception of kinematics, e. g., [DKSF04], observation of kinematics of virtual arms [PFB+01], and the perception of (partially rigid) avatars, e. g., [VMR+04, SWK+06, DBC+06]. A recent study by Farrer *et al.* involves the rotation of a virtual hand derived from deviations of a joystick [FFG+03]. This study aims at identification of actions caused by oneself, i. e., the experience of agency. The aforementioned studies indicate the importance of simultaneous kinematics for presence. Comparable to the degree of realism, direct kinematic feedback is regarded as an influencing factor that increases presence.

1.3.4 Methods for the Measurement of Usability and Presence

Presence and usability are first and foremost subjective concepts that deal with perception and the processing of information in virtual environments. Their assessment and evaluation is not always clear and obvious due to their subjective background.

The measurement of usability is closely linked to the evaluation of user interfaces. General metrics and evaluation methods are described in [HH93, Chapter 10] and for the more sophisticated case of 3D interfaces in [BKLP05, Chapter 11]. Bowman *et al.* emphasize system performance, task performance, and user preference metrics as characteristic factors of user interfaces that would eventually lead to conclusions on system usability (*cf.* Fig. 1.6).

Several methods are proposed for the measurement of presence [IdRFA00, SvdSKvdM01]. Subjective, as well as objective measures are discussed in literature and remain a topic of research. Nearly all studies involve questionnaires designed to measure the subjective opinion about the experienced presence. Frequently used questionnaires are the Presence Questionnaire (PQ) by Witmer and Singer (latest version discussed in [WJS05]), the SUS questionnaire by Slater and colleagues (described in [UCAS00]), and the IPQ system by Regenbrecht and Schubert [RS02] (original set of questions is in German). A factor analysis has been carried out on both the PQ and the IPQ, highlighting significant factors for the measurement of presence (*cf.* section 1.3.1.2). Although questionnaires are a useful means for measuring presence, their exclusive use for reaching conclusions has been questioned and further supportive measures are suggested for inclusion in presence evaluation [Sla04], e. g., physiological measurements.

Metrics based on physiology are regarded as an objective approach to presence measurement. The utilization of physiological reactions in order to measure mental states dates back to Carl Gustav Jung (*cf.* [SA03]). It is obvious that physiological measures can only be obtained in environments demanding for (even unconscious) reactions, e. g., stressful environments. Meehan *et al.* record heart rate, skin temperature, and skin conductance in order to measure presence [MIWB01, MIWB02]. In particular, a correlation is discussed between presence (measured through the SUS questionnaire) and skin conductance, as well as heart rate. The authors discuss reliability, validity, sensitivity, and objectivity of these physiological measures and conclude that heart rate, and to a certain extent skin conductance, satisfy the aforementioned requirements, and can, therefore, be used to measure presence. Comparable results are reported in [WDW98]. In this particular study increased skin conductance was measured in environments with higher subjective presence ratings. Physiological measures are thus a promising means for objectively assessing presence. However, the results have to be carefully analyzed. As Schuemie *et al.* points out, physiological reactions are essentially linked to arousal [SvdSKvdM01], and interpretations may, therefore, be confounded by the simple excitement of the user about the technology being used.

1.4 Discussion of Implications of Presence

The concept of presence, defined as the subjective sensation to be within a mediated environment instead of the surrounding reality, allows an inclusive evaluation of applications and particular realizations of virtual environments. Because presence involves subjective internal and objective external factors it can not be concluded that an environment capable of inducing high sensation of presence perfectly matches the real world. If a user is not interested in the content, presence will be low, no matter how perfect the virtual stimuli are. Accordingly, low presence sensation could even be measured in bland scenarios in reality. Then again, high presence does not necessarily impose a perfect simulation and hardware realization in a particular application or virtual environment. Yet presence is a powerful measure to evaluate variations in application designs and virtual environments and allows for meaningful conclusions because it closely reflects the mental processing of interaction processes in the real world.

So far, the presented concepts and studies have imposed three main topics worth of further evaluation: the general interrelation between presence and usability, the influence of avatars on presence, and methods for measurement of presence. These areas are discussed in the sequel.

As measures of usability are broadly applied to evaluate virtual environments it is of interest to further understand the interrelation between presence and usability. Even though this dependence has not been clarified in each and every detail, a positive impact has been concluded from several studies. Furthermore, it has been shown that this relation also depends on the application itself and on attentional resources involved. Thus, irrespective of the fact that further investigations have to be performed to fully clarify the interrelations between presence and usability, presence is commonly considered to be a highly promising measure for the description of usability characteristics in virtual environments.

Despite the extensive discussions in literature on the overall influence of avatars on presence two open questions remain: first, how is this interrelation expressed in current, sophisticated virtual environments? And, second, what is the contribution of simultaneous kinematics? The first topic arises from the fact, that most studies applied HMD-based or desktop virtual environments with often additional limitations to hardware (no or limited tracking, no stereoscopic representation, *etc.*), and that only basic virtual humanoid methods have been employed. No functional or behavioral study on avatars and presence has been reported that applied fully interactive virtual humanoids in combination with a realistic representation of the avatar in current, room-mounted virtual environments. Reasons for this situation are not obvious and may lie in the highly sophisticated VR hardware and the complexity of the simulation of representation and kinematics. The work required for realization of a virtual humanoid that could be used in such a study should not be underestimated. The described virtual humanoid toolkits in section 1.1.1 could principally be applied in order to enable realistic avatars, but have not been used as avatars for

presence evaluations in room-mounted virtual environments so far. Furthermore, they impose specific bindings to VR software or are not flexible enough to allow for modifications of the simulation of kinematics as discussed in section 1.2. Another reason can be derived from the fact that room-mounted virtual environments do not restrict vision to only computer generated stimuli as in HMD scenarios. The user is aware of and can see his own body throughout the mediated experience. This fact, i. e., the observation of the own avatar while perceiving one's real body, may lead to additional questions requiring answers: the essential open issue focuses on the fact that in a room-mounted set-up the user has to identify with two bodies on spatially shifted positions—his natural real body, and the virtual body of the avatar—whereas equipped with an HMD only the virtual body of the avatar is visible to the user. Thus, it can be concluded that additional research in room-mounted virtual environments regarding avatars and presence is imperative. The second topic addresses the main factor of kinematics of the avatar that seems to have the largest impact on the sensation of presence induced by the avatar. It is of large interest to identify factors that contribute most to the sensation of presence. For a designer of virtual environments it would be helpful to know how to model a virtual avatar and where to invest more: in computational resources and algorithms for a better degree of realism, or in kinematic simulation. However, this has not been researched sufficiently so far and needs further clarification.

The last open issue deals with the basic technique of measurement of presence. Altogether, measurement of presence is based on the established, yet controversially discussed method of questionnaires. First attempts to measure presence with objective physiological measures have been performed. However, additional data are necessary to clarify the potential of objective measures to estimate the experience of presence.

In summary, this chapter has presented research on presence and focused the discussion on the capability of avatars to alter the sensation of presence. In addition, the concepts of presence and usability have been reviewed. As discussed before, both interrelations are currently not clarified. Even though it has been assumed that an application of virtual humanoids as avatars in virtual environments is gainful for both. presence and usability, further studies are imperative to reveal the factors contributing to these complex phenomena.

2
Motivation and Scientific Goals

The previous discussion on present simulation approaches to virtual humanoids (*cf.* section 1.2) and on presence (*cf.* section 1.4) have convincingly demonstrated several deficiencies in current methods and concepts. Specifically, the development of data structures and algorithms for interactive virtual humanoids, as well as investigations into the basics of the influence of avatars on users were found reasonable to be further addressed in order to contribute to a thorough understanding and, hence, reliable and predictable knowledge about the application of virtual humanoids in virtual environments. These considerations resulted in the following main goals:

- Methods for a comprehensive and interactive simulation of virtual humanoids, integrated in an open, easily adaptable framework.
- Tools for virtual humanoids enabling further research on interrelations between users and avatars in virtual environments.
- Experimental evaluation of the influence of avatars on users in virtual environments focusing on presence and kinematic simulation.

These goals are addressed in two parts: first, the elaboration of software solutions capable of realizing the required virtual humanoid methods, and second, the use of these technologies for evaluation of the experimental issues in virtual environments. They imply the following technical topics:

- Design of a data structure being able to hold multimodal humanoid data, and capable of representing the different modalities for representation, as well as the manifold structures of the human anatomy.

- Development of an extensible and flexible system enabling a broad range of simulations of virtual humanoids.
- Solutions for the interactivity requirement of virtual environments even for time-consuming simulations of virtual humanoids.
- Development of multimodal, immersive representational techniques, and, most important, the synchronization across modalities.
- Realization of a robust, whole-body tracking system for direct user–avatar interaction enabling a continuous animation even in tracking environments with sensor drop-outs.
- Analysis and evaluation of methods for motion synthesis based on pre-recorded motion data.
- Realization of simulation approaches for internal human structures, such as musculature. The simulation should achieve real-time performance and deliver data for musculature representation with a high degree of realism in the appearance and animation.

This virtual humanoid technology is a precondition for a further experimental evaluation of the interrelations between avatars and presence. In addition, this evaluation in virtual environments calls for a well-defined, adaptable, and reliable experimental environment as used in current computer-based neuropsychological experiments. Therefore, the development of software for neuropsychological experiments for specific use in virtual environments appeared imperative. The evaluation itself comprises the following goals:

- Experimental evaluation on embodiment and identification with body representations (avatars) in room-mounted virtual environments. It is of particular interest if results predicted and partially evaluated in HMD-scenarios can be reproduced in room-mounted virtual environments.
- Determination of the influence of synchronous kinematics on the sensation of presence. Does synchronous kinematics lead to a stronger identification with the virtual body representation?
- Evaluation of the influence of presence on usability, i. e., study of the correlation of presence and usability measures.
- Examination of the coexisting subjective and objective measures for presence.

Solutions to the technical goals are presented in chapter 3. The methods for virtual humanoids described are combined in the VRZula toolkit consisting of a flexible data structure for humanoid data within a processing framework, and a

hierarchy of simulation methods. Algorithms are detailed in the areas of multimodal representation, kinematics and physiology. The virtual humanoid toolkit and the integrated methods are evaluated with regard to flexibility and interactivity.

Chapter 4 presents the two studies addressing the experimental goals listed above. The experiments are detailed after an introduction of the hypotheses and a description of the experimental software NeuroMan. The functional imaging study analyzes the effects on users caused by synchronous kinematics of avatars. The second study is focused on measurement of presence and performance in a room-mounted virtual environment. The subjects are immersed in a sports task and confronted with a body-tracked avatar in a third-person perspective. Implications from both studies are discussed in view of advantages of an application of avatars in virtual environments.

3

VRZula – Towards a Flexible and Interactive System for Virtual Humanoids

As discussed in the Introduction (*cf.* section 1.2), there is a need for an extendible system for virtual humanoids capable of including a broad range of anatomical data structures. Such a system should allow for extensions towards a comprehensive simulation of the human body (*cf.* chapter 2). Accordingly, this chapter describes the proposed approach and evaluates its performance. One contribution of this work is the software design of the virtual humanoid toolkit VRZula, consisting of a flexible data structure, and clearly defined data access and modification rules. The main contributions are new algorithms for the synchronization of multimodal representation, strategies for motion synthesis, robust extensions to whole-body tracking, and a new approach to simulation of musculature all of which are described in the following sections.

3.1 Data Structures and Processing

As the simulation of humans is still far from being comprehensive, most approaches focus on selected anatomical areas or physiological systems. Furthermore, internal processes are reduced to simplified abstract models. This current state prevails mainly due to the complexity of the human body and the partially incomplete knowledge even in the sciences directly concerned, namely medicine, biology, and biochemistry. With increasing computational power and new findings on the pro-

cesses in the human organism simulation algorithms have been further developed. However, the parts that are not changing are the anatomical components of the simulated body. As a consequence, the basic data must be separated from the simulation itself. Yet this leads to the question how humanoid data can be accessed and modified so that concurrent algorithms do not induce an inconsistent state of the underlying structure. A second issue deals with distributed virtual environments. In addition to methods for synchronization of remote avatars, distributed algorithmic approaches are needed to guarantee interactivity for sophisticated simulation processes. These issues and the solution proposed are discussed in the following.

3.1.1 Layers as Functional Entities

As described in section 1.1.1, most tools use a structure with a skeleton and a skin layer, some also include a layer with muscles. This thesis proposes to generalize this idea to arbitrary layers that are closely bound to anatomical functional systems. Layers can either be used to distinguish sensory information for multimodal representation or to structure functional and anatomical parts of the humanoid. Each layer describes a consistent *functional entity* that corresponds to a physiological or anatomical body system [Ber02, Sal06]. Another characteristic of layers is the *modality* of the data contained. Layers are specified to hold data for a particular modality.

From a systematic point of view several layers are identified, i. e., the skeletal, muscular, nervous, cardiovascular, respiratory, and integumentary (skin) layer. The following paragraphs describe a basic configuration containing the previously mentioned skeletal, skin, and musculature layers that have been developed and integrated into VRZula.

Skeleton Layer. The skeletal system provides physical support in living organisms and builds the basic structure for locomotion. A typical adult human has 206 bones [Gra05]. The skeleton layer contains an abstract representation of the locomotion of the body, as well as references to corresponding anatomical structures, mainly the bones. As for the other layers, the anatomy is described as surface geometries represented through triangular meshes. This information is mainly used for representational purposes. More detailed data can be used for additional simulations, like physiological soft-tissue simulation or even simulations on the cell-level. The abstract representation of the skeletal system corresponds to the widely spread H-Anim standard [H-A04]. This standard defines a kinematic structure as a tree of *joints*. Each joint has a local transformation M_i. The global transformation $\tilde{M}_i$ is computed as

$$\tilde{M}_i = \tilde{M}_j \cdot M_i^0 \cdot M_i \tag{3.1}$$

where $\tilde{M}_j$ is the global coordinate system of the parent joint and M_i^0 an initial constant transformation. The joint tree defines the whole kinematic structure and can be manipulated with forward kinematics algorithms. Joints may have additional

children called *segments*. The purpose of segments is the link to the previously mentioned anatomical bone representations. The transformation of segments directly depends on the parent's coordinate system. Furthermore, segments build a container for *sites*. These points of interest are located within the coordinate system of the adjoining segment. They represent attachment points for end effectors in inverse kinematics computations, or for musculature simulation.

In humanoid locomotion the local transformation matrix M_i can be approximated as a rotation, only[1], that depends on specified rotational restrictions, i. e., the degrees of freedom (DOFs) of a joint. The whole posture of a skeleton is, therefore, described as a set of quaternions $\mathbf{q} = (q_0, \ldots, q_n)$ for all joints in the skeleton. However, one translation, the position of the humanoid root $\mathbf{v}$, is included, as well. The time-dependent posture of a skeleton is, therefore, defined as

$$P_{\mathbf{q}}(t) = (\mathbf{v}(t), q_0(t), \ldots, q_n(t))$$

for a time value t. In addition, $P_{\mathbf{0}} = \mathbf{0}$ defines the dresspose given by the not necessarily purely rotational initial transformations M_i^0 as referred to in equation (3.1). Furthermore, $P_{\mathbf{q}}(0)$ defines the first keyframe in a sequence. Both initial values need not to be identical.

The skeleton layer contains data necessary for kinematic and dynamic simulation, as well as data for multimodal representation of the skeletal system.

Musculature Layer. This layer represents the data of the muscular system. In combination with the skeleton it has the function of support, locomotion, and maintenance of shape. Muscles are controlled through the nervous system. The musculature layer contains an abstract representation of muscles, as well as links to anatomical representations. Muscles are described as *action lines* that are attached to elements of the skeleton layer. The abstract muscle i is described through local transformations M_j^i of its attachment points that are linked to a site k_j of the skeleton layer. Given the global transformation matrix $\tilde{M}_{k_j}$ of this site the global transformation of attachment point j is given as

$$\tilde{M}_j^i = \tilde{M}_{k_j} \cdot M_j^i \,.$$

In addition, each action line consists of a proximal and a distal tendon segment and the intermediate muscle tissue itself. This distinction is, in particular, important for the computation of the resulting forces. Each abstract muscle contains further data for simulation of dynamics. In order to represent the anatomy each abstract object is linked to a geometrical representation of the particular muscle.

[1] An approximation through rotations is sufficient for the majority of cases. Representations for more complex joints are discussed in literature, in particular, for the shoulder [DRBZ01] or the head–neck system [LT06].

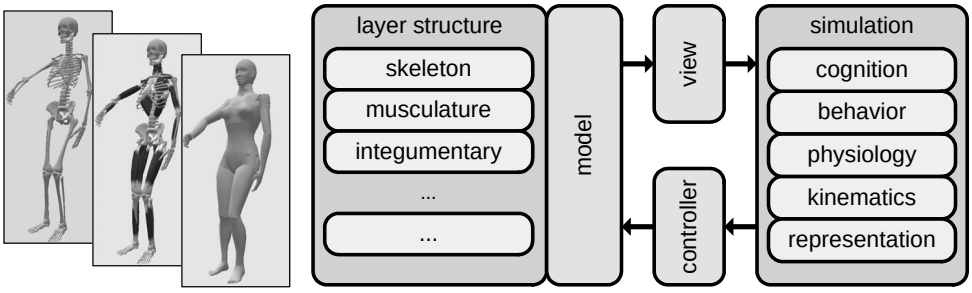


Figure 3.1. The model-view-controller pattern applied on the virtual humanoid data structure: a simulation algorithm reads data of a layer through a specific view and uses a controller to write back modifications.

Integumentary Layer. The skin is with a surface area of 1–2 m^2 the largest organ in the body representing 12–15 % of its weight. Even hair and nails are structures derived from the skin. Skin is continuous with mucous membranes and consists anatomically of two layers. The integumentary system has multiple roles, including physical protection, various sensory tasks, temperature regulation, biochemical synthesis, and absorption. The integumentary layer contains the abstract data for simulation of the broad skin functionality. It builds the interface between physical events in the simulated environment. *Sensory areas* are defined that connect the skin to internal structures. In addition, the skin delivers reactions of the humanoid system that are a result of internal simulations, mainly of kinematic and force simulation in the skeletal and muscular layers. *Deformer* structures are defined for these purposes allowing for modification of the skin. The anatomical representation of the integumentary layer is the skin surface geometry. The geometry needs not to contain skin data, only. In addition, it may hold cloth geometry, hair, nails, teeth, eyes, and the representation of superficial mucus membranes. Alternatively, these structures might be represented by self-contained layers if specific simulations are intended to be used.

Specialization of Functional Entities. The skeleton, musculature, and skin layers have been specified and implemented in the data structure, as well as multiple specializations of the skin layer focusing on facial animation for emotion processing and multimodal speech representation. In the latter case a specialization of the integumentary layer is defined as an audio-visual functional entity that holds data for the visual representation of speech, i. e., the mapping of phonemes to face expressions, and for synthesis and rendering of audio files of the spoken language.

3.1.2 Access and Modification of the Data

In order to control data flow within the virtual humanoid toolkit this thesis proposes to use an updated model-view-controller (MVC) architectural pattern [BMRS96].

This pattern separates the data structure, data interface, and control logic into three components in that modifications to one component can be made with little change to the others.

In the original pattern the view component is described as the user interface, e. g., a GUI displaying the data. In VRZula this concept is used in a more abstract way defining views as the access interface for the model: data of each layer are accessed through a corresponding view. The views are, therefore, not regarded as interfaces for the user, or as a final visualization, but as interfaces for the simulation algorithms. Furthermore, the views have the capability of indicating, which components have been changed since the data was last accessed by a controller. Algorithms can, therefore, identify modified data and induce an update of the simulation.

Controllers are more separated from the views than in the original definition[2]. They represent distinct interfaces that have the capability of modifying the model data. Comparable to the views each controller is associated with one layer. Affected layers are locked during access in order to avoid concurrent writing by simulation algorithms. In addition, flags are updated that indicate local changes to views. Controllers are specialized for particular layers and implement low level functionality to maintain data consistency in the model, e. g., a controller for the skeleton layer implements forward kinematics and joint limit checks.

In our approach each simulation algorithm requests one or multiple instances of views on the layers in the model to keep track of current changes and to access the data. The algorithm, furthermore, requests one controller per layer to modify the model (*cf.* Fig. 3.1). This use of MVC enables a distinct control of data flow and maintains the consistency of layer data for local simulation algorithms. In addition, it allows for a distributed use of virtual humanoids as described in the next section.

3.1.3 Distribution

As mentioned before, distribution is a central concept because of remote avatar synchronization and handling of time intensive simulations. Both aspects can be dealt with by application of the MVC approach. The separation of data and simulation algorithms allows for a distribution of the virtual humanoid on the data level. A pair consisting of a view and a controller is created and serves as the communication channel, e. g., on top of TCP/IP. Upon initialization the view transmits the local data structure. While updated, current changes to the local model are pushed on the communication channel. This information is received by a controller and written to the remote model. The user may freely configure instances of remote simulation algorithms for the distributed model.

A remote distribution of the virtual humanoid as a whole is achieved as follows: the local virtual humanoid is created and all required simulations are instantiated.

[2]In the original description by Buschmann *et al.* the controller responds to events, e. g., user interactions and is, therefore, often implemented as a part of the GUI that, however, represents the view. The functionality of controller and view is, therefore, often combined within one structure.

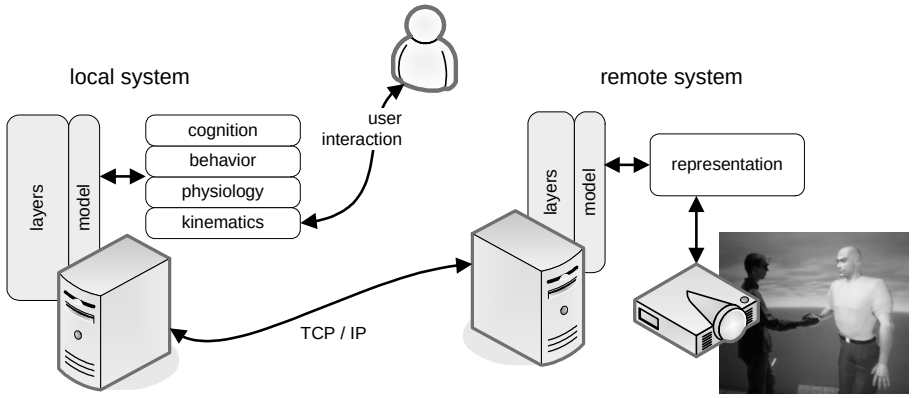


Figure 3.2. Distribution of the data model as used for synchronization of remote avatars and for distribution of simulation algorithms. The TCP/IP communication is realized as a view/controller-pair on the local and the remote site.

In addition, a communication view is linked to the model. The remote application creates an empty humanoid data structure, as well as views and simulation algorithms for representation of the model. A communication controller that connects to the local view is instantiated. The current data model is transferred after both communication parts have been successfully connected. The local user may, e. g., modify the skeleton layer through kinematic simulation. The updates are then transferred to the remote site and displayed accordingly. This basic synchronization and visualization of the remote avatar is depicted in Fig. 3.2.

The previously described method uses a passive remote representation of the virtual humanoid, i. e., no modifications to the model are done on the remote site. In addition, the distribution on the data level allows for the use of simulated remote instances. Consequently, algorithms that would slow down the representation can be distributed to remote instances and only changes are propagated to the local virtual humanoid. The process is as follows: after creation of the local virtual humanoid selected local simulation algorithms are added so that the system remains interactive. A remote representation is created and connected to the local model. Additional simulation algorithms are added on the remote site and simulation results are transferred back to the local model through an additional view/controller-pair. Time-stamps are used in order to avoid circular updates between the remote and the local model.

The distribution approach on the data level simplifies the use of remote virtual humanoids. Once the basic data is synchronized the communication channels are loaded with updates, only. In addition, this approach does not require the same simulations to be run on all instances and allows for a balanced computation.

3.2 Algorithms for the Simulation of Virtual Humanoids

Whilst the previous section has focused on software design of a virtual humanoid data structure and its access control, here the basic structure is extended towards a framework for the simulation of virtual humanoids.

As presented in the Introduction, a large amount of simulation approaches for virtual humanoids is available. However, these algorithms are often utilized in toolkits focused on specific application areas, thus, not allowing for flexible extensions to the simulated data or the algorithms themselves. Only few approaches have been reported that address software architecture issues and that allow for flexible handling of data and algorithms, namely the humanoid component of VHD++ (and its predecessors) and the Virtual Human project (*cf.* section 1.1.1). The basic ideas of these tools are adopted and extended towards a flexible framework for simulation of virtual humanoids. The seamless integration of algorithms within the system proposed is described in order to demonstrate and validate the ideas of this approach including, in particular, novel algorithms in the areas of synchronization of multimodal representations, robustness of body tracking, strategies for motion synthesis for navigation, and musculature simulation.

A modified modeling hierarchy based on Funge's approach presented in section 1.1.2 is used in order to group and integrate the loosely bound simulation algorithms into one system. The hierarchy of five simulation levels is modified to suit the purposes of virtual humanoid simulation and virtual environments. The concept is combined with Kim *et al.*'s form, function, and behavior triad [KKKL98], as well as with the distinction of appearance, animation, and behavior by Thalmann [Tha06]. The first level consists of algorithms for *multimodal representation* and fulfills the tasks of form and appearance. It is derived from the geometric level and based on the fact that multimodality is a key issue in virtual environments. Algorithms on that level typically have access to all layers in the model and display the data for the diverse modalities. The next two levels focus on *kinematics* and the simulation of *physiology* in the virtual humanoid. They correspond to the kinematics and physics levels of the original approach and represent the tasks of function and animation. Kinematic algorithms typically address the skeleton and the musculature layer to compute the locomotion of the virtual humanoid. Physiology algorithms are mainly bound to internal layers and simulate the physiological systems represented by these layers. The last two levels address the *behavior* and *cognition* simulation. Behavior and cognition affect several layers in the model. It is interesting to note though, that there is no direct anatomical layer representing the simulation data for both levels. Behavior and cognition are psychological concepts and the link to anatomical structures has not been clarified (*cf.* also the discussion on the mind-body problem, e. g., [Brü96]).

Table 3.1. Simulation algorithms integrated within VRZula. Methods marked with # are new contributions or contain novel aspects.

Level	Method	
representation	*visualization*	
	• rigid skin and stitching	
	• (GPU-based) vertex blending	
	• blendshapes	
	audio-visual rendering	
	• speech synthesis	
	• audio-visual synchronization	#
kinematics	*forward kinematics*	
	• animation of motion data	
	inverse kinematics	
	• analytical approaches for limbs	
	• Jacobian-based approaches	
	• cyclic coordinate descend	
	body tracking	
	• robust orientation-based tracking	#
	motion synthesis	
	• motion graphs	
	• strategies for construction of motion graphs	#
physiology	*musculature simulation*	
	• inverse dynamics	
	• volume-preserving deformation	#
behavior	*emotion processing*	
cognition	*path planning and navigation*	

The following sections describe simulation algorithms integrated in the VRZula toolkit at the levels of representation, kinematics, and physiology as detailed in Table 3.1.

3.2.1 Multimodal Representation

In computer graphics, algorithms for representation refer to visualization and rendering processes whereas in virtual environments representation aims at displaying data and content for all human modalities, i. e., for the human senses. Therefore, algorithms for multimodal representation should provide display and rendering methods for as many senses as possible. Usually, content is represented on modality-specific displays, e. g., visual representation on screens, auditory represen-

tation through loudspeakers, or haptics through mechanical devices. All displays must be synchronized in order to induce a high degree of immersion.

A representational algorithm is focused on the display of one layer of the model. Upon instantiation the algorithm connects to a view of the layer and retrieves the modality-specific data, e. g., geometries or text for speech representation. In addition, the algorithm connects to a view of the skeleton layer in order to retrieve spatial information and posture details. Data from additional layers can be integrated, as well. In order to enable level-of-detail techniques multiple representations are created on demand, and activated and deactivated on-the-fly. This is also useful, for instance, for interactive modeling purposes in virtual environments. The VRZula toolkit contains algorithms for visualization and speech representation that are described in the following sections.

3.2.1.1 Visualization

Visualization is the classical approach to representation of content. Most VR toolkits are capable of rendering some geometrical description of an object in an immersive way. A virtual humanoid toolkit focused on application in virtual environments must offer a range of visualization algorithms allowing for trade-offs between realism of appearance and rendering performance. Several visualization algorithms discussed in section 1.1.3 have been integrated into VRZula: *rigid skin*, *stitching*, *vertex blending*, and the *blendshape* approach.

Rigid skin visualization uses geometric primitives or predefined geometries. The skeleton layer, i. e., joints, their connectivity, segments and sites are visualized with this approach (*cf.* Fig. 3.3 *a*). The skeletal structure is mapped on scene graph nodes updated according to the current joint transformation. The nodes hold the initial geometries for representation. This approach is also chosen for the visualization of the skin and the musculature layer. Skin is visualized with rigid geometries that represent particular body parts (*cf.* Fig. 3.3 *b*), whereas musculature is visualized either with cylinders to represent action lines or as ellipsoids resembling the muscle distribution. Methods for musculature visualization through arbitrary volume-preserving geometries have been also added. The deformation of these geometries is discussed in section 3.2.3. The final representation is done with rigid skin methods.

Stitching is used to eliminate visual artifacts (cracks and discontinuities on the surface) of the rigid skin approach. Neighbor geometries are searched for contours that are connected with triangle strips. The geometries are visualized with the scene graph-based approach of the rigid skin. The triangle strip geometry is recomputed in each frame and visualized independently. This imposes an additional computational complexity. However, the approach closes the disturbing artifacts at a low rendering cost (*cf.* Fig. 3.3 *b*, *c*). Smooth representation of the human body is achieved with vertex blending. The method uses one geometry and weight factors as defined in the integumentary layer. The toolkit contains a software implementation, as well as an adaptation to graphics hardware. After preprocessing the GPU-based algorithm

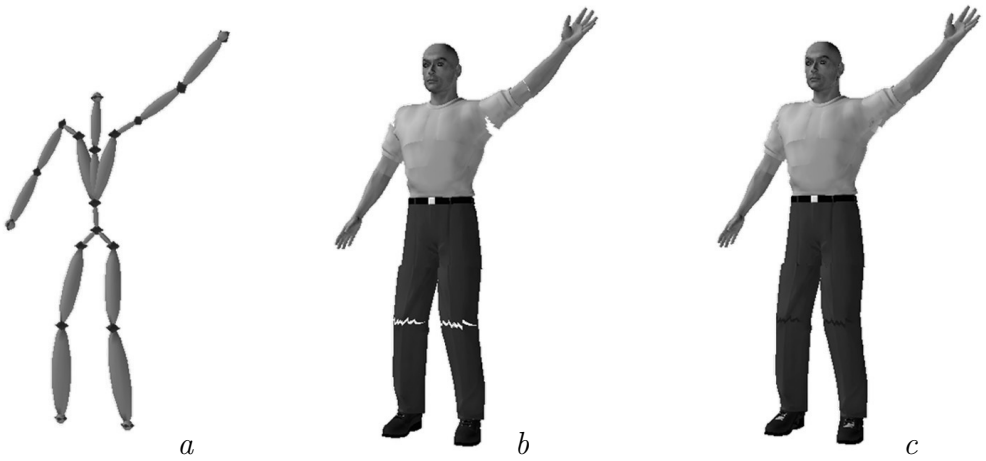


Figure 3.3. Visualization: (*a*) and (*b*) rigid skin visualization, and (*c*) the stitching extension.

transfers the geometry and the weights to the graphics adapter. During interaction only the corresponding joint orientations are updated. The implementation basically follows the elementary vertex blending method described in section 1.1.3.2.

The blendshape approach uses a set of example geometries in order to create a final visualization. These geometries are retrieved from a specialization of the skin layer that also stores the current weight values. The algorithm computes the final geometry as a weighted sum of the targets.

3.2.1.2 Speech Synthesis

The ability of a virtual humanoid system to represent speech is essential for its usability and acceptance. The framework integrates a distributed and adaptable approach for speech synthesis and facial animation. The analysis of current text-to-audio-visual-speech (TTAVS) systems led to an identification of two main issues that have been considered in the software design. The first one is related to the performance of the system. Since speech and binaural synthesis, as well as facial animation represent computationally expensive processes, parallel computations must be employed as much as possible. As the integration of third party TTAVS software seems to be imperative open interfaces have to be defined. Thus, interchangeability becomes a second major design criterion.

As most of the existing systems do not provide data about performance and time behavior, the performance of the whole system with regard to its responsiveness has been analyzed. The duration of the audio-visual synthesis process is considered as one part of the time needed by a virtual character to respond to the user's input. In order to make statements about the performance of the system the durations of the different computational steps involved in the synthesis process have been analyzed. Those values are used to identify bottlenecks and to conclude possible improvements

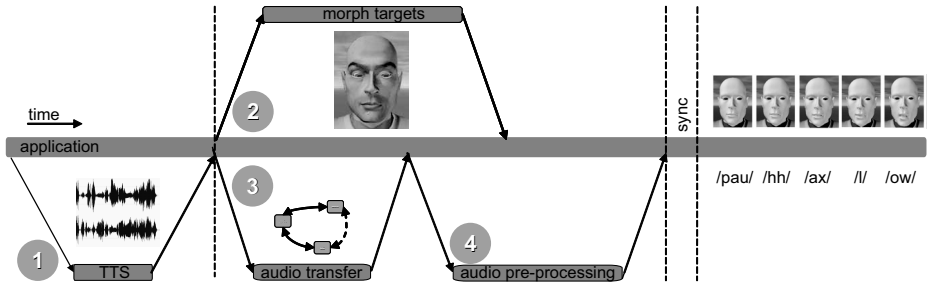


Figure 3.4. Audio-visual synthesis: (*1*) the input text is synthesized, the TTS server returns audio data to the audio server and phonetic data to the client application. (*2*) The synchronization unit picks the facial geometries corresponding to the phonetic data as morph targets. (*3*) The audio data is transferred to the audio server simultaneously. (*4*) Finally, the audio data is loaded into memory of the audio server. After the synchronization phase the animation starts.

that lead to an interactive, multimodal virtual humanoid representation. Results are presented in section 3.3.2.

The distributed approach enables the use of virtually any speech synthesis software that provides data access by means of an API. Because of the focus on the representation of virtual humanoids in immersive audio-visual environments the integration of specialized VR hardware and software plays a key role. Our design approach allows for an integration of further modules like recognition and dialog-generating components to gradually extend the conversational capabilities of a humanoid model.

System Modules. From a modal point of view the audio-visual simulation is divided into an auditory and a visual component. Both parts are connected to a synchronization unit to generate the final audio-visual speech representation. The simulation connects to a speech and a facial layer that contain the simulation data, i. e., the input text that has to be synthesized, and morph targets for facial animation.

The auditory component is responsible for speech synthesis and output of the speech. The synthesis is performed on a text-to-speech (TTS) server that implements a wrapper for third-party speech synthesis software. Virtually any synthesis software that provides access to phonetic information can be attached to the TTS server. The simulation algorithm accesses the server via network, sends input text and receives the generated audio data and phonetic transcriptions. The wrapper of the TTS server transfers the received input text to the synthesis software and sends back the generated results. Comparable to visualization, the audio output is done by the underlying VR toolkit. In some toolkits a dedicated resource is used for audio representation. Therefore, the TTS server can be directly connected to this rendering resource in order to avoid multiple transfers of the audio data.

The visual representation of speech is done through facial animation using the previously described blendshape approach. The system uses morph targets defined in the integumentary layer[3]. Timing data and viseme information from the TTS synthesis are used for generation of the final facial visualization.

Because the audio and the video parts are not aware of their mutual existence both modalities are synchronized. The synchronization unit establishes a link between the auditory and visual simulation part. Several factors, like network, audio and video hardware latencies, as well as the distance of the listener to the speakers, are considered in the synchronization process.

Synthesis Process. The synthesis process consists of several steps (*cf.* Fig. 3.4): after retrieval of the data from the respective layers the input text is synthesized. The TTS server sends the generated audio data to the audio renderer of the VR toolkit and the phonetic data back to the auditory module. While the audio data is transferred and processed by the audio resource, the video module simultaneously prepares the morph targets corresponding to the phonetic transcription. After both processes are completed audio and video outputs are started simultaneously as described.

3.2.2 Kinematic Simulation

Kinematics is the most essential low level simulation task. Kinematics is basically involved in all simulation stages dealing with animation. This is due to the fact that kinematic simulation directly controls the skeleton layer that is central for representation, physiology, and even for behavioral and cognitive tasks. This section describes the methods implemented for the control of kinematics, body tracking, and motion synthesis.

3.2.2.1 Kinematic Control

The control of kinematics is done directly through a controller for the skeleton layer. This controller gives access to joints of the kinematic configuration. Forward kinematics, therefore, directly updates the local rotations $\mathbf{q} = (q_0, \ldots, q_n)^T$ in the desired joints. The global end effector position and orientation is given by the kinematic function $F(\mathbf{q}) = \tilde{M}_e$.

Closed Form Solution for the Human Limb. As discussed in section 1.1.5.1, inverse kinematics solves for the inverse kinematics function $F^{-1}(\tilde{M}_e)$, i. e., the rotations q_i of joints in the kinematic chain involved are determined such that $F(\mathbf{q})$ results in the desired end effector transformation $\tilde{M}_e$. A closed form solution can be derived for short kinematic chains, such as limbs. It is obvious that one DOF

[3] A basic implementation consists of nine visemes for visual speech. Basic emotions, as well as closed eyes and raised eyebrows can be rendered through eight further geometries.

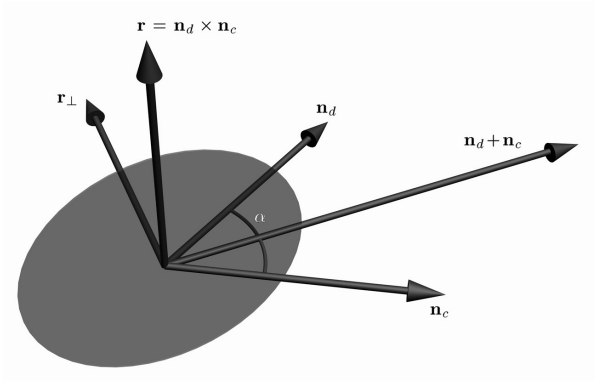


Figure 3.5. Derivation of the quaternion-based CCD approach with positional and orientational goals. $\mathbf{n}_c$ denotes the direction to the current end effector and $\mathbf{n}_d$ is collinear to the desired direction.

remains undetermined, because 6 DOFs are given through the end effector transformation $\tilde{M}_e$ but a human limb contains 7 DOFs: for a human arm the shoulder can be simplified as 3 DOFs, the elbow as one DOF, and 3 more in the wrist[4]. This undetermined DOF can be made clear by imagining a human hand fixed in one location. Without change in the position of the shoulder the arm can still be rotated on an imaginary circle perpendicular to the shoulder–wrist axis. The closed form solution described in [LS99] that is also applicable for simulation of legs has been integrated into VRZula.

Jacobian-based Solution for Arbitrary Kinematic Chains. Analytical solutions exist for chains with a limited amount of DOFs only. Jacobian approaches define a general schema that allows for solutions of arbitrary chains. The solutions differ in their approach to the configuration of the undetermined DOFs, *cf.* section 1.1.5.1 for further details. The following methods have been implemented and are based on the Jacobian matrix J: the transformed Jacobian using $J^T \cdot K$, the pseudo inverse $J^+ = J^T(JJ^T)^{-1}$, and the damped pseudo inverse $J^* = J^T(JJ^T + \lambda I)^{-1}$. The last two approaches can be further extended by including secondary tasks in the null space of the solution. The Jacobian and its inverse representations are computed as described previously.

Quaternion-based Cyclic Coordinate Descend. Jacobian-methods are general approaches, however, they require a large amount of computations. The aforementioned cyclic coordinate descend (CCD) method is derived from geometric conclusions made from the inverse kinematics problem. A kinematic chain is iteratively

[4]This perspective is easier for mathematical computations but does not reflect exactly the biomechanics of the human skeletal system that defines the elbow and the wrist each as 2 DOFs joints. The twist of the wrist is regarded as a DOF of the elbow.

adjusted starting at the end effector joint back to the root joint of the chain. The applied rotation is derived from a geometrical construction. The process is iterated until the desired goal has been reached, or if it turns out that the goal can not be achieved. The CCD method has been implemented and extended insofar as the rotational distance influences the computation of the positional distance. Each joint in the kinematic chain is updated in that the direction to the current end effector $\mathbf{n}_c$ is collinear to the desired direction $\mathbf{n}_d$ (*cf.* Fig. 3.5). This is done by rotation along the axis $\mathbf{r} = \mathbf{n}_d \times \mathbf{n}_c$ with the angle $\alpha = \cos^{-1}(\mathbf{n}_d \cdot \mathbf{n}_c)$, given that both $\mathbf{n}_c$ and $\mathbf{n}_d$ are normal vectors. The choice of the rotational axis is arbitrary. In fact, all axes lying within a plane spanned by the vector $\mathbf{r}_\perp = \mathbf{r} \times (\mathbf{n}_d + \mathbf{n}_c)$ can be used for rotation with an appropriate angle α_i. $\mathbf{r}$ is not defined, or more precisely, equals to zero if $\mathbf{n}_d$ and $\mathbf{n}_c$ are already parallel. In this case, however, the goal of collinearity is already achieved and nothing has to be done.

After rotating with the quaternion $[\alpha, \mathbf{r}]_q$ the direction to the current end effector position $\mathbf{n}_c$ is recomputed. The rotational distance is used to update the twist along this axis. Given the current rotation of the end effector q_c and the desired orientation q_d, a rotation of the current joint with the quaternion $\Delta q = [\phi, \mathbf{s}]_q = q_d \cdot q_c^{-1}$ would result in the desired end effector orientation, however, it would also change the position of the end effector, which is not desirable. Therefore, the rotational axis is restricted to $\mathbf{n}_c$ and the angle β computed that best approximates the original quaternion Δq. This angle can be computed by application of the logarithm on both quaternions and computing the projection of $\log(\Delta q)$ on $\log([0, \mathbf{n}_c]_q)$. The projection $\mathbf{t}$ is computed as follows. Given that $\log(\Delta q) = \log([\phi, \mathbf{s}]_q) = \phi\mathbf{s}$ and $\log([0, \mathbf{n}_c]_q) = \mathbf{n}_c$, as well as the fact that $|\log([0, \mathbf{n}_c]_q)| = |\mathbf{n}_c| = 1$ then $\mathbf{t} = (\phi\mathbf{s} \cdot \mathbf{n}_c) \cdot \mathbf{n}_c$. The second update quaternion, as well as the angle β are obtained by application of the exponential map

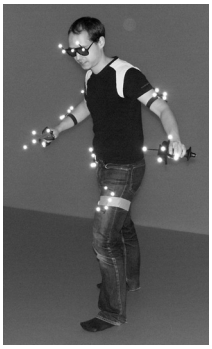
$$\exp(\mathbf{t}) = [\phi\mathbf{s} \cdot \mathbf{n}_c, \mathbf{n}_c]_q = [\beta, \mathbf{n}_c]_q \,.$$

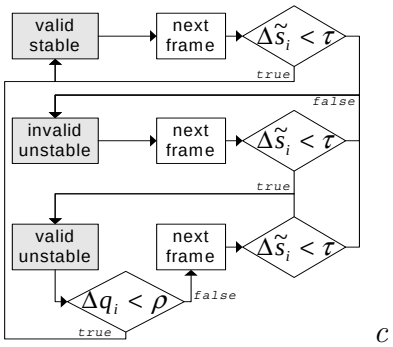
A consecutive rotation with $[\alpha, \mathbf{r}]_q$ succeeded by $[\beta, \mathbf{n}_c]_q$ results in the best approximation of the positional and the orientational goal in each joint.

3.2.2.2 Tracking of Articulated Structures

Beside a direct manipulation of joints or computation of transformations for end effectors, the tracking of articulated objects is essential for interaction in virtual environments. A solution for the tracking of objects with a skeletal configuration based on the work by Molet *et al.* [MBT99] is described along with extensions for a robust tracking with optical sensors. The aim of our adaptation of Molet's method to optical tracking is to enable a short and simple initial phase and to avoid the drawbacks of self-calibration as discussed in section 1.2, like a large amount of tracking sensors, expensive computations imposing latencies, and problematic handling of sensor drop-outs.


a


b



c

Figure 3.6. Orientation-based tracking: (a) congruent dresspose of a tracked virtual humanoid and the initial posture for the tracking sensors, (b) user configured with optical sensors, and (c) states and transitions for detection of valid sensor data. $\Delta\tilde{s}_i$ is the current change in sensor data, and $\Delta q_i = |\log(q_i' \cdot q_i^{-1})|$ the angular distance between the current and the determined joint configuration. τ and ρ are the predefined thresholds.

Orientation-based Tracking. Orientation-based tracking requires that the articulated structure of the object that has to be tracked is known *a priori*, which is the case in most applications. That is, it can be said, whether an arm, or a human, or even an animal has to be tracked. Furthermore, the approach requires the correspondence between the sensors used and the topology of the tracked object.

As this tracking approach is solely based on orientations, the global transformation for a joint i, as given in equation (3.1) on page 50, is rewritten as

$$\tilde{q}_i = \tilde{q}_{i-1} \cdot q_i^0 \cdot q_i \tag{3.2}$$

where $\tilde{q}_i$ is the global orientation, q_i^0 the dresspose orientation, and q_i the current local orientation of joint i. $\tilde{q}_{i-1}$ is the global orientation of the parent joint $i-1$. Correspondingly, the global dresspose orientation of a joint i is given as

$$\tilde{q}_i^0 = \tilde{q}_{i-1}^0 \cdot q_i^0 \,.$$

It can be easily shown that if $\forall j \in \{0, \ldots, i\} : q_j = \mathbf{0}$ then $\tilde{q}_i^0 = \tilde{q}_i$.

Furthermore, the algorithm uses the current global orientation $\tilde{s}_i$ of the tracking sensors i. Without loss of generality, it is assumed, that sensor i corresponds to joint i. In order to map the current sensor orientation on the associated joint it is necessary to record an initial dresspose orientation $\tilde{s}_i^0$ of the tracking sensors. The method requires that the dresspose of the tracked object and the initial posture of the sensors are congruent as demonstrated in Fig. 3.6 *a*.

Each current orientation of a sensor $\tilde{s}_i$ is interpreted as a deviation Δs_i from the recorded dresspose

$$\tilde{s}_i = \tilde{s}_i^0 \cdot \Delta s_i \,,$$

which immediately yields $\Delta s_i = (\tilde{s}_i^0)^{-1} \cdot \tilde{s}_i$. In analogy, each local orientation of a joint in the tracked kinematic structure can be expressed as a deviation Δq_i from

the dresspose

$$\tilde{q}_i = \tilde{q}_i^0 \cdot \Delta q_i \,.$$

It has to be noted, that Δq_i is the deviation from the initial dresspose and must not be confused with the local orientation q_i as defined in equation (3.2) that—in contrast—relies on the global transformation $\tilde{q}_{i-1} \cdot q_i^0 \neq \tilde{q}_i^0$. However, if the initial dresspose of the kinematic structure $\tilde{\mathbf{q}}^0$ and of the tracking sensors $\tilde{\mathbf{s}}^0$ are congruent, the equality of both deviations can be assumed, as well, i. e., $\Delta q_i = \Delta s_i$. This leads to the following relation

$$\begin{aligned} \tilde{q}_i &= \tilde{q}_i^0 \cdot \Delta q_i \\ &= \tilde{q}_i^0 \cdot (\tilde{s}_i^0)^{-1} \cdot \tilde{s}_i \,. \end{aligned}$$

In order to determine the local orientation q_i of the joint i that corresponds to the global orientation $\tilde{q}_i$ all previous joints $j \in \{0, \ldots, i-1\}$ in the kinematic chain leading to joint i have to be adjusted beforehand. When this is the case, i. e., the global orientation $\tilde{q}_{i-1}$ of the parent joint $i-1$ is up-to-date then equation (3.2) can be transformed to

$$\begin{aligned} q_i &= (q_i^0)^{-1} \cdot (\tilde{q}_{i-1})^{-1} \cdot \tilde{q}_i \\ &= (q_i^0)^{-1} \cdot (\tilde{q}_{i-1})^{-1} \cdot \tilde{q}_i^0 \cdot (\tilde{s}_i^0)^{-1} \cdot \tilde{s}_i \end{aligned}$$

and yields the current local orientation q_i that is used to update the current joint i.

Robustness. The described method delivers stable results if sensor data are delivered continuously. Temporal sensor failures are, for example, caused by occlusion of optical markers, or ferric interferences with electromagnetic sensors. These errors are, in general, handled by the tracking hardware and the sensors are marked as invalid, however, they result in a discontinuous stream of sensor data, which is, furthermore, often biased directly before or after a drop-out of a sensor. As this behavior results in disturbing jumps in the kinematics of the avatar, the algorithm is extended in that sensor data are tested on-line for validity. In addition, the algorithm allows for the use of redundant markers.

In a first step the raw data are checked for validity. For this, let $\tilde{s}_i$ be the current sensor data, $\tilde{s}_i^\lambda$ the last valid sensor data, t the current time stamp, t^λ the time stamp when the last valid sensor data has been obtained, and q_i the current local joint configuration. The idea is to compute the new joint orientation based on a comparison of the rotational distance between the last and the current orientation with time-dependent thresholds. The change in current sensor data

$$\Delta \tilde{s}_i = |\log(\tilde{s}_i \cdot (\tilde{s}_i^\lambda)^{-1})|$$

is compared with the threshold

$$\tau = \tau_0 + (t - t^\lambda) \cdot \Delta\tau$$

that determines the maximal rotational difference assumed to be valid. If $\Delta\tilde{s}_i \geq \tau$, then the sensor i is marked as *invalid* and *unstable* and q_i is left unchanged in this frame. The time-dependent increase of the threshold guarantees that the system is not trapped in a deadlock situation. However, if the threshold τ has not been exceeded, then $\tilde{s}_i^\lambda$ and t^λ are updated to the current values and a candidate for the current local joint configuration q_i' is computed as discussed. Yet if the sensor has been marked as *unstable* previously, then q_i' is adjusted in that it does not exceed a maximal rotational angle ρ. Therefore, if $|\log(q_i' \cdot q_i^{-1})| \geq \rho$, then

$$\alpha = \frac{\rho}{|\log(q_i' \cdot q_i^{-1})|}$$

and the joint orientation is computed as the slerp interpolation (*cf.* [DKL98]) between q_i and q_i', namely $\text{slerp}(q_i, q_i', \alpha)$. Otherwise, the orientation is directly set to q_i' and the sensor is marked as *stable*. The three states and their transitions are shown in Fig. 3.6 *c*. If a sensor is in a state other than *valid/stable*, then the algorithm checks for redundant sensors and considers those data for computations, instead.

The described algorithm represents a robust adaptation of Molet's technique for real-time tracking systems that often accept temporal drop-outs of sensors. It features a fast don-session, as well as a short calibration procedure. The results of this technique are presented in section 3.3.3.

3.2.2.3 Motion Animation and Synthesis

The VRZula toolkit contains approaches to motion animation and synthesis. An animation is defined as a sequence of keyframes determined by postures $P_{\mathbf{q}}(t) = (\mathbf{v}(t), q_0(t), \ldots, q_n(t))$. Data for a time step between two keyframes are computed through interpolation, e. g., linear or spline interpolation for translational data, and slerp interpolation for orientations. The interpolated data is used as input for kinematic algorithms described in section 3.2.2.1.

Animation Data. VRZula is capable of loading motion animation data. This data is preprocessed to meet the requirements of the skeleton layer of the model. This is done by motion retargeting methods described in the Introduction. The method proposed here does not assume a topological identity of the origin and the target skeletal structure, but requires a mapping of corresponding joints in each skeleton. The dresspose transformation of these joints is updated in that the subsequent offset directions are collinear in the target skeleton (*cf.* Fig. 3.7). Given the source direction $\mathbf{n}_s$ and the dresspose of the target $\mathbf{n}_t$, collinearity is achieved through a rotation along $\mathbf{n}_s \times \mathbf{n}_t$ and a subsequent twist around $\mathbf{n}_t$ as derived for the CCD approach on page 61. If the subsequent twist is not defined in the source file, then a heuristic-based twist is assumed.

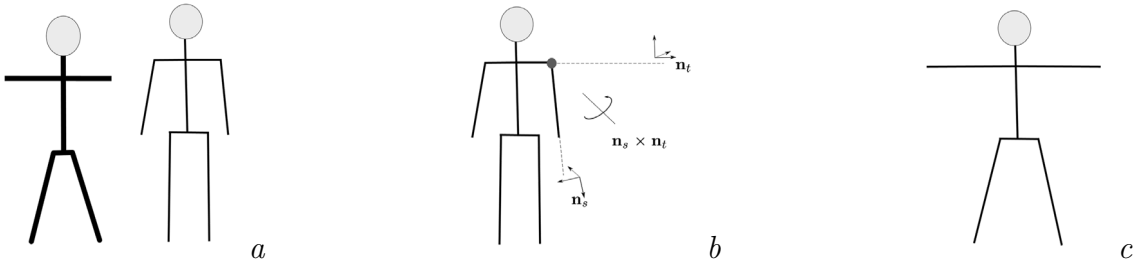


Figure 3.7. Retargeting procedure: (a) the dresspose of the target (left) and the source (right) skeleton, (b) the source skeleton is adjusted insofar as its offsets are collinear with the source skeleton, and (c) the final retargeted source skeleton.

Retargeting. The retargeted skeleton is animated with the original animation data, but at least one additional step that takes the size of the animated figure into consideration is required. In this step the root position is updated with the aim of minimizing footskating. Spacetime constraints and their time intensive computation are not applied. Instead, the distance between the humanoid pelvis and the feet for both, the source l and the target skeleton l' is used. This distance is well approximated as the sum of lengths of the thigh and the lower leg. The ratio l/l' between the length of both offsets is used to scale the root position. This approach assumes that motion is performed on planar ground. If terrain information is included, scaling has to be done relative to the elevation. In other words, the terrain elevation defines the $x \times y$ plane, which is regarded as the origin for all scaling.

Synthesis. If a desired motion is not included in the original set of animation sequences $\Pi = \{P_i(t)\}$, it could be created by manipulation of end effectors and application of inverse kinematics methods. However, this approach is difficult, time intensive and often does not deliver motion resembling natural kinematics, typically due to the amount of undetermined DOFs of the skeleton. A more promising method is to combine parts of existing pre-recorded motion in that it approximates the desired motion (*cf.* section 1.1.5.2). The motion graph approach has been integrated in the system proposed. The original ideas have been extended by applying improved transitions between two motion sequences and by using multiple strategies for motion graph construction.

Improved Transition Generation. A transition $P_T(t)$ of duration Δt is defined between time steps of two motions $P_1(t_1)$ and $P_2(t_2)$. In order to avoid undefined keyframes[5] at the end of P_1 and in the beginning of P_2 the motion is extended on both ends such that $P_1(t) = P_1(t_1^{\max})$ if $t > t_1^{\max}$ and $P_2(t) = P_2(0)$ if $t < 0$. With

[5]Both cases can occur if $t_1 + \Delta t > t_1^{\max}$, or if $t_2 - \Delta t < 0$.

an interpolation function $\alpha(t) \in [0;1]$ that meets the additional requirements

$$\alpha(t) = \begin{cases} 0, & t \leq 0, \\ 1, & t \geq 1. \end{cases}$$

a transition is defined as

$$P_T(t) = (1 - \alpha(t)) \cdot P_1(t + t_1) + \alpha(t) \cdot P_2\left(t + t_2 - \Delta t\right) . \tag{3.3}$$

Basically each interpolation function delivers a valid transition, however, continuous functions, like the linear interpolation $\alpha_{\text{lin}}(t) = {}^{t}/_{\Delta t}$, deliver more meaningful results. Better results are obtained by interpolation functions whose first derivatives are continuous on their boundaries, i. e., $\dot{\alpha}(t) = 0$ for $t \in \{0, \Delta t\}$. This condition is satisfied, for example, by

$$\alpha_{\text{cub}}(t) = -2\Big(\frac{t}{\Delta t}\Big)^3 + 3\Big(\frac{t}{\Delta t}\Big)^2 .$$

The major drawback of this classical interpolation approach is, however, that feet will move on the ground without pelvis motion, hence, the classical footskate problem appears in this context. This thesis proposes to avoid footskating by restricting start and end keyframes for transitions to the moments where one foot is planted on the ground (and thus can be fixed) and the other lifted (*cf.* Fig. 3.8). The position of the root of the kinematic structure is modified in that the planted foot remains in the same position throughout the whole transition. To find a suitable area the neighborhood in the time domain of postures in the original motion $P_1(t_1')$ where $t_1' \in [t_1 - \varepsilon; t_1 + \varepsilon]$ and the target motion $P_2(t_2')$, $t_2' \in [t_2 - \varepsilon; t_2 + \varepsilon]$ are analyzed. A transition is rejected if no keyframes with the given condition in a certain neighborhood are found. If suitable keyframes are found, then t_1 and t_2 are adjusted to t_1' and t_2', respectively.

$\phi(P)$ is defined as the global position of the planted foot for a posture P. The root position of the target motion is aligned such that

$$\phi(P_1(t_1)) = \phi(P_2(t_2)) . \tag{3.4}$$

However, this alignment does not imply that the considered foot remains in the same location during transition. This is achieved by the following additional transformation.

The transition $P_T(t) = (\mathbf{v}(t), q_1(t), \ldots, q_n(t))$, $0 \leq t \leq \Delta t$ between $P_1(t_1)$ and the modified $P_2(t_2)$ is computed according to equation (3.3). Furthermore, the difference of the foot position during the transition $P_T(t)$ is given as

$$\Delta\phi(t) = \phi(P_T(0)) - \phi(P_T(t)) .$$

Here, this difference is applied to the root position in order to achieve a stable planting of the foot. The footskate-clean transition is given as

$$P_T'(t) = P_T(t) + \Delta\phi(t) = (\mathbf{v}(t) + \Delta\phi(t), q_1(t), \ldots, q_n(t)) .$$

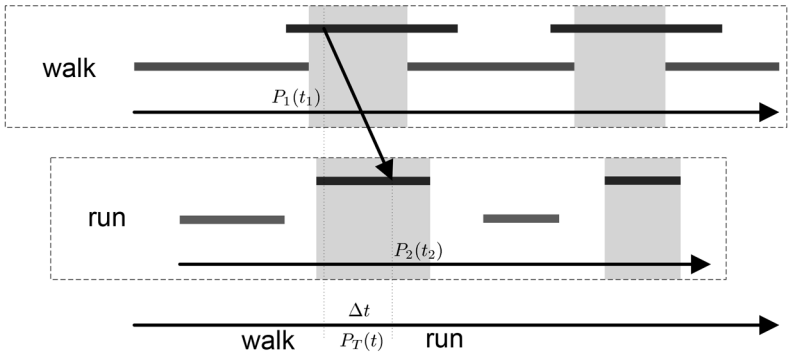


Figure 3.8. Footskate-clean transitions. The figure shows two motion sequences P_1 (walk) and P_2 (run) and the time intervals with a planted right (red) or left (blue) foot. Transitions are searched for between the gray areas, only, where one foot is planted and the other lifted. The transition is processed insofar as the planted foot remains in one location.

The following proof shows that the new transition $P'_T(t)$ preserves the continuity of the original transition $P_T(t)$.

Proof. In order to show the continuity of $P'_T(t)$, it is sufficient to show that $\Delta\phi(t)$ is a function with the required continuity, and that it equals to zero on definition boundaries.

$\Delta\phi(t)$ is a sum of functions whose continuity depends on the definition of $P_T(t)$, in particular, $\alpha(t)$. Thus, the sum of these functions is continuous, as well. Given $P_T(0) = P_1(t_1)$ and $P_T(\Delta t) = P_2(t_2)$ the value of $\Delta\phi(t)$ on its boundaries 0 and Δt is computed as

$$\Delta\phi(0) = \phi(P_T(0)) - \phi(P_T(0)) = 0\,,$$

and

$$\Delta\phi(\Delta t) = \phi(P_T(0)) - \phi(P_T(\Delta t)) = \phi(P_1(t_1)) - \phi(P_2(t_2)) \overset{(3.4)}{=} 0\,,$$

therefore, $\Delta\phi(0) = \Delta\phi(\Delta t) = 0$. With the continuity preserving characteristic of $\Delta\phi(t)$ and the boundary property follows that $P'_T(t)$ preserves the continuity characteristic, as well. □

The presented method for transition generation uses a general blending approach between two motion sequences and extends it in a way that results in footskate-clean transitions. The concatenation of motion sequences from the set Π that allows for synthesis of numerous new motions is described in next paragraph.

Motion Graph Construction. Motion graph techniques have been described in section 1.1.5.2. The following basic approach has been included in VRZula along with the presented method for transition generation. In addition, several strategies for determination of transitions within the motion graph are proposed.

A motion graph is built from a set of original motions sequences $\Pi = \{P_i(t_i) \,|\, 0 \leq t_i \leq t_i^{\max},\ i = 0, \ldots, N\}$ that are retargeted to one skeleton. These sequences are taken as input, included in the motion graph, and connected by transitions. Finally, the graph is cleaned up by searching for the largest strong component and by pruning the rest.

Transitions are included based on a quality measure that compares two postures $P = (\mathbf{v}, q_0, \ldots, q_n)$ and $P' = (\mathbf{v}', q_0', \ldots, q_n')$ of one skeletal structure as defined in [WB03] and given as

$$|P, P'| = |\mathbf{v} - \mathbf{v}'| + \sum_i w_i \,|\log(q_i'^{-1} \cdot q_i)| \tag{3.5}$$

where w_i are normalized weights that reflect the particular importance of a joint. A set of estimates is given in [WB03]. Such a measure reflects well the similarity of a posture, however it can not measure similarity of dynamics in the origin and target motion. Therefore, a measure that compares multiple postures during the intended transition has been used. Given two motions $P_1, P_2 \in \Pi$ and two time steps t_1, t_2 for the transition candidates $P_1(t_1)$ and $P_2(t_2)$, as well as the transition duration Δt. The time aligned postures during, before, and after transition, i. e., the measures $|P_1(t_i + t_1), P_2(t_i + t_2 - \Delta t)|$ are computed for a set of samples t_i. These measures are combined in the sum

$$\Theta(T) = \sum_{t_i} u_i \,|P_1(t_i + t_1), P_2(t_i + t_2 - \Delta t)| \tag{3.6}$$

with normalized weights u_i. A set of samples and weights that compares the start and end pose, as well as one intermediate value is given as $(t_0, t_1, t_2) = (0, \Delta t/2, \Delta t)$ and $(u_0, u_1, u_2) = (2/5, 1/5, 2/5)$. Two transition candidates are connected if their similarity measure is below a certain threshold $\Theta(T) < \theta^{\max}$.

Suitable pairs $P_1(t_1)$ and $P_2(t_2)$ for transitions are found based on the following techniques:

Complete search strategy S_0: The straightforward approach as used by several methods discussed in section 1.1.5.2 analyzes the whole data base, i. e., it computes the measure $\Theta(T)$ for each pair of keyframes in all motion sequences, and searches for local minima to find the optimal set of transitions.

While strategy S_0 requires extensive computations even for mid-size data bases and, therefore, is not applicable for fast creation of motion graphs, the following set of strategies is proposed that optimize the motion graph for navigation purposes:

Random strategy S_R: A previously defined number of transitions is searched for in a random way. For this purpose two motion sequences $P_1, P_2 \in \Pi$, and two time steps t_1, t_2 with $0 \leq t_1 \leq t_1^{\max}$ and $0 \leq t_2 \leq t_2^{\max}$ are determined and define a transition T. The transition is discarded if $\Theta(T) > \theta^{\max}$, or included otherwise. The search process is continued until the designated amount of

transitions has been found, or if an upper limit for total searches has been exceeded.

Cycle strategy S_C: This strategy is based on the fact that most motions have a cyclic nature. Furthermore, motions like walking that can be re-executed periodically have an enlarged coverage and wider applicability. For this reason each original motion is searched for suitable transitions from the end to the beginning of the sequence. The algorithm analyzes a predetermined number of frames in the beginning $\{t_1^0, \ldots, t_1^n\}$ and frames in the end $\{t_2^0, \ldots, t_2^n\}$ of a sequence $P(t)$ and chooses the best pair $P(t_1^i)$, $P(t_2^j)$ in terms of the measure $\Theta(T)$.

Direction strategy S_D: For navigation purposes the direction of a motion is important for the coverage of an area. This thesis proposes to use a classifier that analyzes motion in terms of change in direction of the pelvis. All original motions in the motion graph are divided into segments that start and stop with a planted foot; then, the difference in pelvis orientation is computed. A segment of the motion P_j is thus given through the starting posture $P_j(t)$ at time stamp t. The classifier now assigns each segment to a directional class C, such that a set of motion segments $\Pi_C = \{P_{i_0}(t_0), \ldots, P_{i_m}(t_m)\}$ with $P_{i_j} \in \Pi$ is obtained satisfying the classifier's condition. In order to connect motions with varying properties, transition candidates are searched for as follows: for each pair of classifiers $C \neq C'$, two representative motion segments $P_i(t_i) \in \Pi_C$, $P_j(t_j) \in \Pi_{C'}$ are chosen at random and the best suitable transition is included in the graph. Intuitively, this procedure finds interesting transition points helping to connect motions with different styles; the achieved connectivity of the motion graph allows for a synthesis of motion streams of a larger variety.

After determination of transitions, i. e., the insertion of new edges into the motion graph, the global structure is optimized. Similar to [KGP02], proximal transitions representing similar information are searched for and, in addition, the largest strong component of the graph is identified. The rest of the graph is pruned. The loss of this information is justified because unreachable parts, or *cul-de-sac* transitions are not useful for animation and navigation purposes. Fig. 3.9 depicts a motion graph, its strong components and the pruned rest, however, typical motion graphs have a much larger amount of nodes and edges.

Application of the Motion Graph for Animation. The motion graph is used to animate the skeleton layer of a virtual humanoid. Each path in the graph results in a smooth animation that has a high visual plausibility. As mentioned in the beginning of this section, the graph is built with the purpose to animate the kinematics of the virtual humanoid in a way that is not covered by the original set of motions Π. The formulation of this desired motion is transformed into a graph search and the resulting path in the graph yields the corresponding animation. The A* algorithm

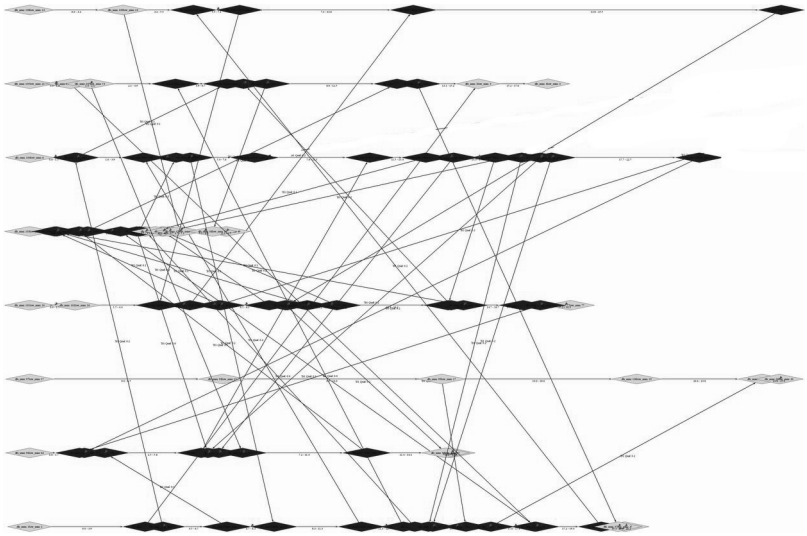


Figure 3.9. Motion graph constructed from eight initial motion sequences. Pruned parts are marked gray. Nodes colored in blue demarque the largest strong component of the motion graph.

is applied for all motion search tasks. The resulting path is described as a sequence of edges $S = e_0, \ldots, e_m$. The particular algorithm searches the graph with the aim to find an optimal sequence. The following descriptions of a desired motion and corresponding algorithms give an overview of the flexibility of this approach.

Navigation. The task is to find a sequence that transforms the virtual humanoid from the current place to a desired position and orientation. As each animation sequence results in a new global position and orientation, the graph search algorithm has to find a path that yields a minimal transformation error. Given the sequence S, its final transformation $M = [\mathbf{v}, q]^T$, and the desired transformation $M_d = [\mathbf{v}_d, q_d]^T$. The task is formulated as the minimization of the metric

$$\Xi_\Delta(S) = \xi \cdot |\mathbf{v}_d - \mathbf{v}| + (1 - \xi) \cdot |\log\left(q_d \cdot q^{-1}\right)|$$

where $\xi \in [0, 1]$ is a weight determining the importance of the position and the orientational measure. A search is performed in the graph to find the sequence with an optimal Ξ_Δ. Because the spatial coverage of a one-pass search algorithm is limited the graph search is repeated until a desired goal distance has been reached or no improvement can be achieved. Alternatively, the graph is searched in multiple passes. This corresponds to a multiple unfolding of the graph in the spatial domain.

Pose. The animation task is to reach a specific pose P_d from a given start pose, e. g., the current animation position. The graph is searched for the node with minimal posture deviation as defined in equation (3.5). This node is found by

a breadth-first search in the graph. In addition, a threshold θ is defined and all nodes having a better metric than θ are included in a set of candidates. After determination of the target set a graph search is performed that computes the optimal path from the current node.

Style. The task is to perform a motion defined by a category of style C. As laid out before, categories may be described as the direction of motion, or as arbitrary motion styles, e. g., running, walking, dancing, etc. Given the current animation state in the graph, either an edge or a node, the task is to reach an arbitrary edge in a set E' that represents the desired style. The query is defined as a set of nodes P' given as the start nodes for each edge $e \in E'$. P' is thereafter used in the *pose* approach.

The graph search for the optimal path depends mainly on two metrics, the quality and the duration of the motion. The created motion should have a high *quality*, i. e., the amount of transitions should be limited. This is formulated by using the measure $\Theta(T)$ defined in equation (3.6) that has been computed during the creation of the motion graph. The metric $\Theta(e_i)$ for an edge e_i, with $i \geq 1$, is defined as the quality of the transition between e_{i-1} and e_i. The measure for the whole sequence is given as

$$\Xi_Q(S) = \sum_{i=1}^{m} \Theta(e_i)\,.$$

The task of the graph search is to find a sequence with a minimal cost $\Xi_Q(S)$. The *duration* of a motion is important for the description of the pace of reactivity. It is defined as

$$\Xi_D(S) = \sum_{i=0}^{m} |e_i|$$

with $|e|$ being the duration of the animation of the edge e. The task of the search algorithm is to find an S with minimal $\Xi_D(S)$. Both measures, Ξ_Q and Ξ_D, are combined and used in the A* search to determine the optimal path S.

This section has presented approaches for kinematic simulation that have been seamlessly integrated in the multi-level simulation structure of VRZula. The main contribution in this area are extensions to orientation-based tracking that increase the robustness of the original approach by delivering continuous data even during sensor drop-outs. Performance data and results are presented in section 3.3.3. Motion synthesis has been handled in great detail including preprocessing steps and synthesis with motion graphs. Strategies for the automatic construction of motion graphs, as well as the application of the motion graph for several tasks have been described. The consequences of the strategy chosen on navigation capabilities are presented in section 3.3.4. Methods for kinematic simulation presented in this section are the fundament for the experiments described in chapter 4.

3.2.3 Physiology and Physical Simulation: Muscle Simulation and Deformation

Physiology affects the whole human body. This simulation level subsumes a wide range of approaches that describe the physical properties of various body systems. Simulation of nerves, for example, affects the nervous system and has an impact on the majority of other systems. Simulation of forces during locomotion is more specific and mainly affects the musculoskeletal system, i. e., the skeleton and attached muscles. This section describes simulation algorithms for physiology on the basis of a specific approach to musculature simulation.

Simulation of musculature and its effects on the skeletal system are prominent examples in physiology. The main reactions of the human body are mediated through musculature, e. g., locomotion, movement of limbs, eyes, and other body parts, as well as speaking is dependent on some action of musculature. Furthermore, muscle simulation involves two system layers, in particular, the skeleton and the musculature layer, and is being affected by results from the neural simulation.

This section presents a fast approach to musculature simulation that involves inverse dynamics from the skeleton layer in order to compute muscle forces. A second issue of simulation is the deformation of muscles during motion. The aim is a realistic representation of muscles. Current approaches follow the path of finite elements simulation and often do not achieve interactive frame rates. The proposed simulation and deformation method uses volume-preserving scaling factors as described in subsequent sections.

3.2.3.1 Muscle Simulation

Muscles are represented as connecting parts between attachment points on components of the skeleton layer. This corresponds to the action line approach described in section 1.1.6, where an action line is defined through its insertion and origin attached to a bone.

The biomechanical simulation for each muscle is solved through inverse dynamics. The forces are simulated through a general linear spring-mass model. The algorithm uses the current skeletal configuration and computes the force for each muscle affected by the kinematics of the skeleton. Given the length l of a relaxed muscle, its current length l', and the resulting expansion difference $\Delta l = l' - l$. The relative force parallel to the action line is approximated as

$$F_{\text{rel}} = k_s \cdot \Delta l + k_d \cdot \Delta v \,.$$

Typical mean values for the spring constant k_s for skeletal musculature are discussed in [Pag02]. The damping part is usually neglected, i. e., $k_d = 0$. In order to compute the absolute force produced by the muscle the current length l' is taken into consideration. This is done by more detailed approaches described in the Introduction,

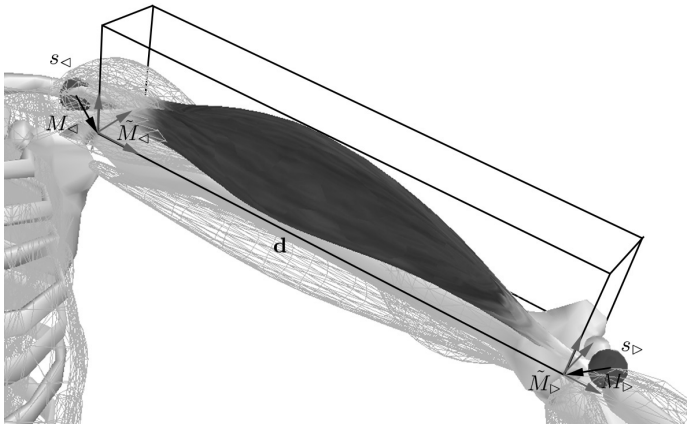


Figure 3.10. An arbitrary muscle geometry enclosed by a bounding box used in the volume-preserving approach. $\tilde{M}_\triangleleft$ and $\tilde{M}_\triangleright$ are the global insertion and origin transformations marked as the green coordinate systems. $M_\triangleleft$ defines the transformation of the muscle with respect to the insertion site $s_\triangleleft$. The origin vector **d** leads to a constant local transformation $M_\triangleright$ of the origin with respect to the attachment site $s_\triangleright$. The muscle is deformed depending on the current transformation of both attachment sites (marked as red spheres).

e. g., methods proposed by Hill or Zajac (*cf.* section 1.1.6.1). The handling of relative forces allows for fast computations and a comparison of force-characteristics in muscles.

3.2.3.2 Volume-Preserving Muscle Deformation

As described in section 3.1.1, an abstract muscle is defined through attachment points, its local coordinate systems M_j^i, and the forces taking effect in the particular segment of the muscle. The musculature data is usually represented by geometric primitives, like lines, narrow cylinders and ellipsoids. This representation allows for a schematic overview of the musculature structure, however, it lacks realistic appearance. The approach presented proposes to use arbitrary geometries as muscle representations that are created from anatomical reference data. These geometries are deformed in that they preserve the volume. The method proposed is distinct in such a way that it allows for realistic appearance and interactive frame rates. Furthermore, it does not require exhaustive computations, like in current finite element approaches. In fact, the results are directly used as scaling factors in the scene graph and no further modification of the dresspose geometry is required.

Algorithm. Each action line segment of the abstract muscle is assigned to an arbitrary geometry. The geometry has to be defined with the insertion point of the muscle located in the origin of its local coordinate system. Furthermore, the local coordinate system should be defined with its axes collinear to the scaling directions.

The origin of the muscle, finally, is defined through a translation along the x-axis, given by the vector $\mathbf{d} = [x_{\triangleright}, 0, 0]^T$, or the corresponding translation matrix $M(\mathbf{d})$, respectively. If no origin offset is provided, the maximum extension of the muscle on the x-axis $x^{\max}$ is taken to define the origin vector as $\mathbf{d} = [x^{\max}, 0, 0]^T$.

The muscle geometry is attached in the dresspose of the skeleton, whereby the insertion and origin of a muscle are attached to two sites $s_{\triangleleft}$ and $s_{\triangleright}$, respectively. The muscle geometry coordinate system is linked to the global coordinate system $\tilde{M}^0_{s_{\triangleleft}}$ of the insertion site through a local translation $\mathbf{v}^0_{\triangleleft}$ and rotation $q^0_{\triangleleft}$ that aligns the muscle with the skeleton in its initial dresspose. Thus, the initial global transformation of the insertion (and the whole muscle geometry) is given as $\tilde{M}^0_{\triangleleft} = \tilde{M}^0_{s_{\triangleleft}} \cdot M(\mathbf{v}^0_{\triangleleft}) \cdot M(q^0_{\triangleleft})$. Similarly, the global transformation of the origin in dresspose is computed as $\tilde{M}^0_{\triangleright} = \tilde{M}^0_{\triangleleft} \cdot M(\mathbf{d})$ and the local transformation that associates the origin with the site $s_{\triangleright}$ in the dresspose as $M^0_{\triangleright} = (\tilde{M}^0_{s_{\triangleright}})^{-1} \cdot \tilde{M}^0_{\triangleright}$ with $\tilde{M}^0_{s_{\triangleright}}$ being the global transformation of the origin site $s_{\triangleright}$ in dresspose. $M^0_{\triangleright}$ is used to derive the initial translation $\mathbf{v}^0_{\triangleright}$ and rotation $q^0_{\triangleright}$ such that $M^0_{\triangleright} = M(\mathbf{v}^0_{\triangleright}) \cdot M(q^0_{\triangleright})$. The whole configuration is depicted in Fig. 3.10.

The deformation of the musculature geometry is dependent on the current global transformations $\tilde{M}_{s_{\triangleleft}}$ and $\tilde{M}_{s_{\triangleright}}$ of the attachment sites. Expressed in the coordinate system of the insertion site $s_{\triangleleft}$, the desired transformation of the muscle origin

$$M^d_{\triangleright} = \tilde{M}_{s_{\triangleright}} \cdot M^0_{\triangleright} \cdot \tilde{M}^{-1}_{s_{\triangleleft}}$$

yields the desired position $\mathbf{v}^d_{\triangleright}$ and orientation $q^d_{\triangleright}$ such that $M^d_{\triangleright} = M(\mathbf{v}^d_{\triangleright}) \cdot M(q^d_{\triangleright})$. The task is to find the best quaternion $q_{\triangleleft}$, which aligns the muscle with the desired direction $\mathbf{d}^d = \mathbf{v}^d_{\triangleright} - \mathbf{v}^0_{\triangleleft}$, and which approximates best the two constraint orientations $q^0_{\triangleleft}$ for the insertion and $q^d_{\triangleright}$ for the origin.

To achieve this goal the muscle is rotated along the axis $\mathbf{r} = \mathbf{d}^d \times \mathbf{d}$ with the angle

$$\alpha = \cos^{-1}\left(\frac{\mathbf{d}^d \cdot \mathbf{d}}{|\mathbf{d}^d| \cdot |\mathbf{d}|}\right).$$

Thus, with $\mathbf{n_r}$ being the normalized axis $\mathbf{r}$ the rotation is done through the quaternion $[\alpha, \mathbf{n_r}]$. To best approximate the constraint orientations the muscle is subsequently twisted along the new direction $\mathbf{d}^d$ that is collinear with the local x-axis after rotation. The logarithm notation for quaternions is used to compute the projection of $q^0_{\triangleleft}$ and $q^d_{\triangleright}$ on the rotational axis of $\mathbf{d}$, i. e., $\mathbf{n}_x = [1, 0, 0]^T$. Given a weighting factor $\xi \in [0; 1]$ the projection of $\log(q^0_{\triangleleft}) = \log([\gamma_{\triangleleft}, \mathbf{n}_{\triangleleft}]_q) = \gamma_{\triangleleft}\mathbf{n}_{\triangleleft}$ and $\log(q^d_{\triangleright}) = \log([\gamma_{\triangleright}, \mathbf{n}_{\triangleright}]_q) = \gamma_{\triangleright}\mathbf{n}_{\triangleright}$ is computed as

$$\log([\beta_\xi, \mathbf{n}_x]_q) = \beta_\xi \mathbf{n}_x = \big(\xi \cdot (\gamma_{\triangleleft}\mathbf{n}_{\triangleleft} \cdot \mathbf{n}_x) + (1 - \xi) \cdot (\gamma_{\triangleright}\mathbf{n}_{\triangleright} \cdot \mathbf{n}_x)\big) \cdot \mathbf{n}_x .$$

Thus, the final musculature rotation is given as

$$q_{\triangleleft} = [\beta_\xi, \mathbf{n}_x]_q \cdot [\alpha, \mathbf{n_r}]_q$$

and the overall transformation as $M_{\triangleleft} = M(\mathbf{v}^0_{\triangleleft}) \cdot M(q_{\triangleleft}) \cdot M(q^0_{\triangleleft})$. The transformation $M_{\triangleleft}$ has to be recomputed each time the global transformation of the origin attachment site has changed. The application of $M_{\triangleleft}$ results in geometries aligned with the current posture. Deformation of the muscle geometry is described in following.

Computation of Scaling Factors. The algorithm described in the last section transforms the musculature geometries in that they are aligned with the current skeletal posture. However, the muscle has to be deformed to fit in the current posture. The goal is, furthermore, to deform musculature insofar as the volume remains the same. The volume of a muscle is defined as the volume enclosed by its triangular mesh[6]. This thesis proposes to use scaling factors derived from the bounding box of the geometry.

Given an arbitrary closed triangular mesh $G = (P, T)$ as a set of vertices $P = \left\{\mathbf{v}_i = [x_i, y_i, z_i]^T \mid i \in 0, \ldots, n\right\}$ defined in a local coordinate system and connecting triangles $T = \{\mathbf{t}_{i,j,k} = (\mathbf{v}_i, \mathbf{v}_j, \mathbf{v}_k) \mid \mathbf{t}_{i,j,k} \in P^3\}$, a bounding box is defined as a box, whose axes are aligned to the local coordinate system, and which encloses the geometry. The bounding box is given by two vectors $\mathbf{b}_{\min} = [v_x, v_y, v_z]^T$ and $\mathbf{b}_{\max} = [w_x, w_y, w_z]^T$ where

$$\forall \mathbf{v} \in P: \quad \mathbf{b}_{\min} \preceq \mathbf{v} \preceq \mathbf{b}_{\max}\,.$$

The relation $\preceq$ compares the components of two vectors, i. e., $\mathbf{a} \preceq \mathbf{b} \iff \forall i : a_i \leq b_i$. An alternative description of a bounding box is given by only one corner $\mathbf{b}_{\min} = [v_x, v_y, v_z]^T$ and the size of the bounding box $\mathbf{l} = [l_x, l_y, l_z]^T = \mathbf{b}_{\max} - \mathbf{b}_{\min}$. The components of $\mathbf{l}$ are denoted as length, height and width. Without loss of generality, it is presumed that $\mathbf{b}_{\min}$ is identical to the origin of the local coordinate system.

The algorithm is based on the fact that if a bounding box is scaled, then the volume of the enclosed geometry will scale by the same factor. In particular, this is also true for non-uniform scaling factors. Thus, if volume-preserving scaling of geometries has to be done, it is sufficient to compute volume-preserving scaling factors for the associated bounding box, and to use these values for the scaling of the enclosed geometry.

Volume-preserving scaling factors of bounding boxes are computed as follows. Given the original size $\mathbf{l} = [l_x, l_y, l_z]^T$ of the bounding box, its volume $V = l_x \cdot l_y \cdot l_z$, and the modified length l'_x; then l'_y and l'_z are computed as

$$l'_y = {}^1\!/\rho \cdot \sqrt{V/l'_x}\,, \qquad l'_z = \rho \cdot \sqrt{V/l'_x} = \rho^2 \cdot l'_y\,.$$

The control factor ρ is dependent on the initial ratio between width and height and can be used to influence the expansion of the bounding box if changed. The width-height ratio is given by $\phi = l_z/l_y = \rho^2$. It can be easily verified that $l'_x \cdot l'_y \cdot l'_z = V$,

[6] In order to enable a meaningful definition of volumes the musculature geometries should be defined as closed triangular meshes.

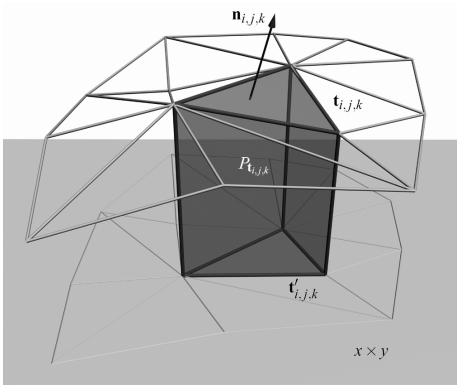


Figure 3.11. Regular prism $P_{\mathbf{t}_{i,j,k}}$ used for computation of the volume of a closed triangular mesh.

i. e., both volumes of the original and scaled bounding box are equal. The volume-preserving scaling factors are given as

$$s_x = l'_x/l_x\,, \qquad s_y = l'_y/l_y\,, \qquad s_z = l'_z/l_z\,. \tag{3.7}$$

Obviously,

$$s_x \cdot s_y \cdot s_z = 1\,. \tag{3.8}$$

Using these factors, the scaled geometry $G^* = (P^*, T^*)$ is computed by scaling the positions of all $\mathbf{v} \in P$, such that $P^* = \{\mathbf{v}_i^* \mid i \in 0, \ldots, n\}$ with $\mathbf{v}_i^* = [s_x \cdot x_i, s_y \cdot y_i, s_z \cdot z_i]^T$. T^* is updated accordingly and includes the scaled triangles. This operation is normally performed through applying a scaling transformation matrix to a geometry node in a scene graph. The correctness of the algorithm is shown in the following proof.

Proof. To prove that the enclosed volume is preserved when the geometry is scaled according to equation (3.7), the computation of the volume of a closed triangular mesh is handled first.

The proof is based on the Divergence Theorem, also known as Gauss' Theorem and discussed for instance in [Bar97]. Given a differentiable vector field $\mathbf{F} : R^3 \mapsto R^3$ in a compact region V enclosed by a surface S then

$$\iiint_V \nabla \cdot \mathbf{F}\; dV = \iint_S \mathbf{F} \cdot \mathbf{n}\; dS$$

where $\nabla\cdot$ is the divergence, $\mathbf{n}$ is the normal vector of the surface boundary S, dV is the infinitesimal volume element, and dS is the infinitesimal element of the enclosing surface area S. By choosing an arbitrary vector field $\mathbf{F}$ with a divergence $\nabla\cdot\mathbf{F} = 1$ the volume integral becomes the volume of the enclosed region, i. e., $\iiint_V \nabla\cdot\mathbf{F}\; dV = V$, resulting in

$$V = \iint_S \mathbf{F} \cdot \mathbf{n}\; dS\,.$$

Because of the fact that the surface S is represented by triangles $\mathbf{t} \in T$ the integral can be rewritten as a Riemann sum

$$V = \sum_{\mathbf{t} \in T} \iint_{\mathbf{t}} \mathbf{F} \cdot \mathbf{n} \, dt \tag{3.9}$$

where dt is the infinitesimal element of triangle $\mathbf{t}$. The remaining task is to calculate $\iint_{\mathbf{t}} \mathbf{F} \cdot \mathbf{n} \, dt$ for each triangle $\mathbf{t} \in T$. Without loss of generality, $\mathbf{F}$ is chosen as $\mathbf{F}_z([x, y, z]^T) = [0, 0, z]^T$. With this choice the triangle integral $\iint_{\mathbf{t}} \mathbf{F}_z \cdot \mathbf{n} \, dt$ describes the signed volume of a regular prism $P_{\mathbf{t}}$ spanned by $\mathbf{t}$ and its projection on the $x \times y$ plane $\mathbf{t}'$, whose sign depends on the orientation of $\mathbf{n}$ (*cf.* Fig. 3.11). As stated above, $\mathbf{F}$ can be chosen freely as long as $\nabla \cdot \mathbf{F} = 1$ holds. For numerical computations of volumes other choices of $\mathbf{F}$ might be more suitable in order to minimize numerical errors. Having defined the vector field $\mathbf{F}$, equation (3.9) is rewritten as

$$V = \sum_{\mathbf{t} \in T} V(P_{\mathbf{t}}) \tag{3.10}$$

where $V(P_{\mathbf{t}})$ is the signed volume of the regular prism $P_{\mathbf{t}}$ that is computed as follows.

Given a triangle $\mathbf{t}_{i,j,k} \in T$, then its normal vector is defined as $\mathbf{n}_{i,j,k} = \mathbf{m}_{i,j,k}/||\mathbf{m}_{i,j,k}||$ with $\mathbf{m}_{i,j,k} = (\mathbf{v}_j - \mathbf{v}_i) \times (\mathbf{v}_k - \mathbf{v}_i)$. In order to compute the volume of a closed triangular mesh it is presumed that all triangles are oriented in a consistent way, e. g., in a counter-clockwise order, causing their normal vectors to point outside of the geometry.

The projection $\mathbf{t}'_{i,j,k}$ of triangle $\mathbf{t}_{i,j,k}$ on the $x \times y$ plane that spans the regular prism $P_{\mathbf{t}_{i,j,k}}$ (*cf.* Fig. 3.11) is given by

$$\mathbf{t}'_{i,j,k} = (\mathbf{v}'_i, \mathbf{v}'_j, \mathbf{v}'_k) = ([x_i, y_i, 0]^T, [x_j, y_j, 0]^T, [x_k, y_k, 0]^T) \, .$$

The lengths of the parallel edges of $P_{\mathbf{t}_{i,j,k}}$ are given by the z-values z_i, z_j, z_k of the vertices of the triangle $\mathbf{t}_{i,j,k}$, whereas the area of the perpendicular cross-section is the same as of the projected triangle $\mathbf{t}'_{i,j,k}$, given by

$$\left| A(\mathbf{t}'_{i,j,k}) \right| = \frac{1}{2} \cdot \left\| (\mathbf{v}'_j - \mathbf{v}'_i) \times (\mathbf{v}'_k - \mathbf{v}'_i) \right\| \, .$$

By calculating the vector product the signed area value is defined that depends on the orientation of the normal vector $\mathbf{n}_{i,j,k}$. This signed area is given by

$$A(\mathbf{t}'_{i,j,k}) = \frac{1}{2} \cdot ((x_j - x_i)(y_k - y_i) - (y_j - y_i)(x_k - x_i)) \, .$$

Therefore, the signed volume of the regular prism $P_{\mathbf{t}_{i,j,k}}$ is computed as

$$V(P_{\mathbf{t}_{i,j,k}}) = A(\mathbf{t}'_{i,j,k}) \cdot \frac{z_i + z_j + z_k}{3} \, .$$

According to equation (3.10) the sum of all signed volumes $V(P_{\mathbf{t}})$ of regular prisms $P_{\mathbf{t}}$, $\mathbf{t} \in T$ results in the volume of the enclosed mesh $G = (P, T)$

$$V(G) = \sum_{\mathbf{t} \in T} V(P_{\mathbf{t}})\,. \tag{3.11}$$

It remains to show that the scaled volume $V(G^*)$ is equal to $V(G)$. For this purpose, it is sufficient to validate that each volume of a regular prism $V(P_{\mathbf{t}_{i,j,k}})$ remains the same after scaling. The scaled signed area is computed as

$$\begin{aligned} A(\mathbf{t}'^*_{i,j,k}) &= \frac{1}{2} \cdot (s_x(x_j - x_i) \cdot s_y(y_k - y_i) - s_y(y_j - y_i) \cdot s_x(x_k - x_i)) \\ &= s_x s_y \cdot A(\mathbf{t}'_{i,j,k}) \end{aligned}$$

and yields directly the scaled volume

$$V(P_{\mathbf{t}^*_{i,j,k}}) = s_x s_y \cdot A(\mathbf{t}'_{i,j,k}) \cdot \frac{s_z(z_i + z_j + z_k)}{3} = s_x s_y s_z \cdot V(P_{\mathbf{t}_{i,j,k}})\,.$$

This results under consideration of equations (3.8) and (3.11) in

$$V(G^*) = \sum_{\mathbf{t}^* \in T^*} V(P_{\mathbf{t}^*}) = \sum_{\mathbf{t} \in T} s_x s_y s_z \cdot V(P_{\mathbf{t}}) = \sum_{\mathbf{t} \in T} V(P_{\mathbf{t}}) = V(G)$$

proving the stated volume-preserving feature for arbitrary triangular meshes. □

This proof has shown that by scaling arbitrary geometries with scaling factors derived from non-uniform volume-preserving scaling of the enclosing bounding box, the volume of the scaled geometry itself remains the same. The algorithm proposed allows for a fast computation of deformation factors for arbitrary musculature geometries and serves as an example for integration of physiological simulations into the VRZula toolkit.

In summary, section 3.2 detailed simulation algorithms integrated in the VRZula toolkit on the levels of representation, kinematics, and physiology (*cf.* Table 3.1). Novel contributions have been presented in the areas of synchronization of multimodal representation, robustness of orientation-based tracking, strategies for motion graph construction, and volume-preserving deformation of musculature. The VRZula toolkit and these contributions are evaluated and discussed in the next section.

3.3 Evaluation

The virtual humanoid data structure, access methods and simulation algorithms proposed have been implemented for the use in virtual environments based on the

VR toolkit ViSTA and its media extensions described in [AKLV04, Ass06]. ViSTA facilitates the creation of VR applications that can be run on a variety of operating systems and VR hardware ranging from desktop-VR to immersive large projection environments. After a presentation of references to applications realized with the VRZula toolkit, the technical contributions of the toolkit are being evaluated in the subsequent sections.

3.3.1 Application of VRZula

The preliminary considerations in the design of the data structure and the processing framework deal with the fact that the basic data affected by the simulation have to be independent from the simulation approach. In order to handle the complexity of the human body a general structure for the respective parts and data objects has been proposed.

The virtual humanoid toolkit VRZula, as well as its data structures and the processing framework proved to be readily applicable in several scenarios. Besides its application in the studies presented in sections 4.3 and 4.4 it has been used in the following areas:

Psychology. The toolkit is integrated in the NeuroMan software frameworks for psychological studies (*cf.* section 4.2). Virtual humanoids are used in the design of experimental set-ups either as avatars or as animated agents. Beck *et al.* use a virtual arm created with VRZula in an fMRI study [BVW+07].

Usability. Virtual humanoids are used for a comparative study on interaction metaphors [VUK05], and subjects had to select and attach example-based muscles on a humanoid skeleton. The humanoid was realized with VRZula in a CAVE-like environment and on a desktop-VR system.

Sign Language. VRZula humanoids are used in experiments on perception of the German Sign Language [UVW+07]. Pre-recorded and preprocessed gestures are animated with a virtual humanoid with changing visual representation and perspective.

Anesthesia. The toolkit is used in the regional anesthesia simulator RASim[7] [UFN+07]. In this work an additional neural layer is integrated and coupled with haptics interaction.

Anatomy. The toolkit is used for representation and simulation of anatomy, in particular, of musculature [VUKB06]. Several simulation approaches were applied on a skeleton and a musculature layer. The distribution of simulation algorithms to remote computing resources ensured interactivity even for time-consuming algorithms.

[7] http://www.rasim.info, last visit May, 2007.

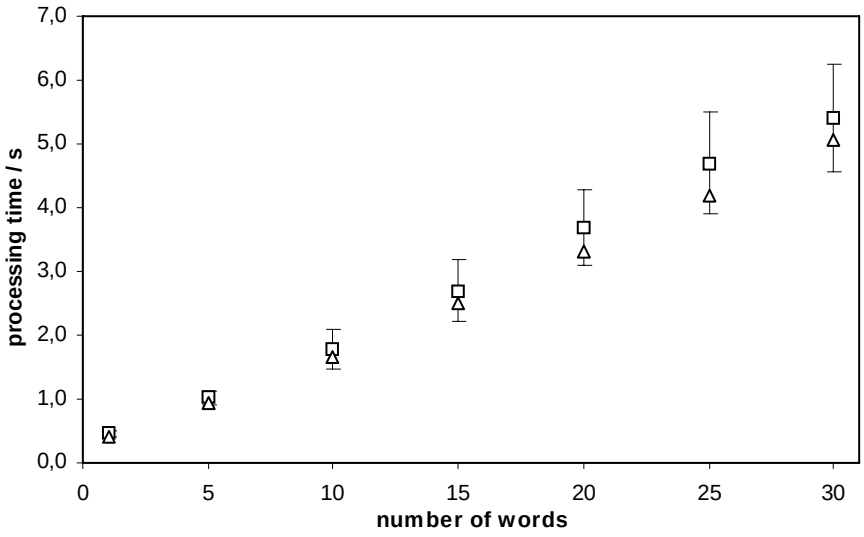


Figure 3.12. Duration and standard deviation of the overall synthesis (□) and the corresponding speech synthesis duration (△).

The described scenarios show that the virtual humanoid toolkit is usable in a broad range of VR applications that demand interactive virtual humanoids. As demonstrated, the data structure proposed is capable of holding comprehensive descriptive data of the human body, like musculature or data representing nerves. Even though the layered approach is inspired by previous methods, it extends the original idea by specifying a layer as a functional entity and clearly distinguishing between data and function. The applications show that this fact makes adaptations and extensions feasible. The proposed processing framework based on the MVC design pattern and its use for distribution represents a clear, yet robust method for the access and modification of data, guaranteeing flexibility with regard to changes on either the humanoid data structure, or the simulation algorithms. This flexibility allows for a wide range of applications and algorithms as presented above.

3.3.2 Audio-visual Speech

The synchronization of auditory and visual output is determined by multiple factors and latencies of the hardware, as well as the network used. These factors have been derived from a performance analysis of the audio-visual component contained in the virtual humanoid toolkit proposed. In order to determine runtime capabilities and response time, the preprocessing and the animation phase have been analyzed, as well. This section describes the set-up of the testing environment and discusses the results.

Table 3.2. Performance of individual synthesis steps for both test series with 20 words per sentence, which is the average sentence length in natural conversation. The step number refers to Fig. 3.4 on page 59. t_{low} denotes the mean duration and the standard deviation for the low-resolution model, and t_{high} for the high-resolution model. The model complexity is significantly involved in the preparation of morph targets.

Process	Step	t_{low} / s	t_{high} / s
speech synthesis	1	3.19 ($\pm$0.56)	3.22 ($\pm$0.52)
network transfer from TTS server	1	0.09 ($\pm$0.01)	0.09 ($\pm$0.01)
morph targets / audio transfer	2/3	0.06 ($\pm$0.01)	0.10 ($\pm$0.00)
preparation of audio data	4	0.27 ($\pm$0.21)	0.29 ($\pm$0.19)
overall process		3.61	3.70

Concerning computation time, it is interesting to observe the behavior of a TTAVS system in the context of a conversation between a user and a virtual human. Even though VRZula is currently not equipped with speech recognition or AI units, the speech generation process is considered as a part of the virtual human's reaction to a verbal interaction with the user. The response time of the system is interpreted as duration of pauses in the so-called turn-taking process, i. e., the switching of the acting role in a conversation. The time needed for turn-taking is, among other things, context sensitive, i. e., it depends on the semantics of the conversation. Weilhammer and Rabold did a statistical analysis on the VERBMOBIL speech corpus, which consists of a large number of spontaneous speech recordings [WR03]. According to their results, an average pause takes about 0.38 s. For the measurements here this value is considered as a lower bound to which the response time of the system should converge. Another important value concerning the behavior of the system is the frame rate at runtime, becoming most critical during display of the animation. Both values have been analyzed using the same hardware set-up with different configurations.

3.3.2.1 Runtime of the Preprocessing Step and Frame Rates

Three Intel-based platforms have been used as components of the system. The client application was run on a 3 GHz Dual Pentium 4 system, the audio server was equipped with an RME Hammerfall audio hardware running on a 3 GHz Pentium 4. A binaural approach has been used for sound rendering as described in [AKL05, LSVA07]. The Festival software (*cf.* section 1.1.4.1) installed on a 1 GHz Dual Pentium 4 architecture has been used for speech synthesis. All peers used TCP/IP and were connected with 100 Mbit Ethernet via 4 hubs with latencies of about 1 ms, except for the connection between audio server and client application, which used the same switch on a 1 GBit network with a latency less than 1 ms. The network adapters were running in full duplex mode.

Computation time during preprocessing and animation is determined by different parameters, like the number of phonemes, the complexity of the model, bandwidth and workload of the utilized network and the computation power of the utilized systems. The duration for each individual part of the synthesis process has been measured. Since the number of phonemes, i. e., the length of the input text, has the biggest impact on the final processing time, the system behavior has been analyzed by synthesizing randomly generated sentences with a length of 1 to 30 words. For each number of words 50 samples were taken. Two different humanoid models have been used in two test series: a low-resolution model consisting of 2,184 vertices and a high-resolution model with 6,284 vertices for the head geometry.

Fig. 3.12 shows the processing time for all test series, Table 3.2 descriptive statistics for the processing of an average sentence. For evaluation of the real-time capabilities of the system the frame rate during speech output has been measured. While the low-resolution model was rendered with an average of 420.0 fps the system was capable of displaying the more complex model at a rate of 64.8 fps.

3.3.2.2 Discussion

The modular design proposed reflects the ideas of some systems discussed in section 1.1.4. Even though distributed systems imply a communication overhead and an increased cost of synchronization, such an approach is advantageous mainly because of a computational speed-up compared to a single platform solution, and the interchangeability of components, like the TTS software.

The performance data clearly indicate a linear relation between the sentence length and the overall processing time (*cf.* Fig. 3.12). The time needed to process an average sentence of 20 words is too long for systems requiring an interactive response time, but then, the data show that speech synthesis done by the exchangeable TTS system requires almost 90 % of the computational time (*cf.* Table 3.2). As expected, the geometrical complexity of morph targets has only an impact on the rendering speed and the duration of the processing of corresponding morph targets (represented by step (2) in Fig. 3.4, page 59). However, it may become notable if models of higher complexity are used. The extensive proportion of the speech synthesis part bears possible improvements. A possibility to speed-up the system is to exchange the TTS system or even to parallelize the TTS stage. Input with multiple sentences could be split up and processed on distributed resources, as well.

The presented TTAVS system enables speech synthesis for virtual humans as one aspect of conversational agents. The approach focuses on the application in virtual environments in order to create a more natural way of human computer interaction by means of immersive visual and auditory speech output. The modular and distributed design has been applied and successfully implemented for the purpose of text-to-speech synthesis and multimodal representation. Some of the performance data obtained indicate bottlenecks, yet also pave the way for improvements to such systems.

Figure 3.13. Body tracking: user movements and the resulting posture of the virtual humanoid.

3.3.3 Body Tracking

The robustness extensions to the orientation-based body tracking in VRZula have been evaluated in a five sided CAVE-like environment using an optical tracking system consisting of four ARTtrack1 cameras (Advanced Realtime Tracking, Weilheim, Germany). The cameras were mounted in the four upper corners of the projection system. The environment was run by altogether 10 clients synchronized by one server. The server was a 3 GHz Dual-Xeon system with 4 GB RAM. The clients were 2.8 GHz Pentium 4 systems with 4 GB RAM and nVidia GeForce 6800GT graphics adapters. The test persons have been dressed with eight sensors in total according to Table 3.3. Each sensor consisted of four or five optical markers. The thresholds for detection of reliable sensor data (*cf.* section 3.2.2.2) have been set to $\tau_0 = 35°$, $\Delta\tau = 0.025°/\mathrm{s}$, and $\rho = 5°$. The virtual humanoid model was an H-Anim-compliant VRZula humanoid with a skeleton layer according to level of articulation one. The skin layer was visualized with 18,537 vertices using GPU-based vertex blending (*cf.* section 3.2.1.1).

The orientation-based tracking system was evaluated by multiple test persons with regard to the usability of the whole procedure. The body tracking was configured in that the upper body was directly tracked and the legs adjusted with inverse kinematics. The test persons have been asked to strike several poses (*cf.* Fig. 3.13) and to judge the resulting posture of the virtual humanoid. The verbal subjective reports have been recorded.

3.3.3.1 Configuration Time and Performance

As the method does not require an accurate positioning of the markers on the body parts, the whole calibration procedure is fast. After placement of the markers, the test persons stood in the same pose as the visualized virtual humanoid in order to

Table 3.3. Placement of sensors and mapping to the skeleton layer of the virtual humanoid. Joint names are given according to the H-Anim notation.

Sensor	Body Part	Joint
1	head	skullbase
2	pelvis (frontal)	sacroiliac
3	pelvis (dorsal)	sacroiliac
4	right upper arm	r_shoulder
5	left upper arm	l_shoulder
6	right forearm	r_elbow
7	left forearm	l_elbow
8	back	vt1

determine the initial sensor configuration. The don-time for eight markers and the calibration has been 2 minutes on average.

The average time for rendering of the virtual humanoid has been 16.50 ms. The algorithm needed 0.08 ms for computation of the posture. The computation time has been lengthened to 0.12 ms with the robustness extensions proposed.

3.3.3.2 Reliability and Validity of the Tracking Data

Altogether 8,424 s of tracking data have been recorded at 60 Hz. In 6.52 % of the time the tracking system reported sensor failure of at least one sensor. Unstable sensors have been detected 5.36 times per minute on average. The correction to $\rho = 5°$ eliminated jumps during the unstable state, however, led to a visual effect that was described as latency of the system by the test persons. The pelvis was tracked with one redundant sensor. Sensor 2 was lost by the tracking system or marked as invalid in 15.7 %, and sensor 3 in 3.4 % of the recorded data. This difference is due to the fact, that sensor 3 was fixed at the back and was, therefore, not occluded during bending of the upper body. Yet both sensors for the sacroiliac joint were invalid on average during 0.4 % of the same time, only.

3.3.3.3 Discussion

The orientation-based body tracking integrated in VRZula has proven to reliably and efficiently determine the posture of a virtual humanoid based on the motion of a user's body. The approach requires a short don and calibration phase because it does not impose exact positioning of the tracking sensors; and delivers motions that closely correspond to those of the user. These advantages are leveled by the necessary assumption that the body of the user and the virtual humanoid are similar in size, and that the posture taken during the calibration phase is identical with the dresspose of the virtual humanoid. Large differences in the proportions of body parts and the corresponding segments of the virtual humanoid lead to inaccuracies

Figure 3.14. Footskate-clean transitions: (*a*) a direct transition contains footskating artifacts (red circle), (*b*) the optimized transition is constructed between frames with one planted foot. The root position is updated during transition in that the planted foot remains in one position on the ground.

in the resulting pose. Yet a precise determination of the user's body size and the adaptations of the skeleton would lead to a system with less usability. The proposed extensions for enhancement of robustness have proven their necessity and quality. The use of even one redundant marker led to a large reduction of invalid data to less than one percent of the total time. The handling of invalid and unstable sensor data eliminated breaks and jumps in the resulting animation of the virtual humanoid that occurred more than 5 times per minute on average.

In summary, the proposed techniques are fast extensions to already established tracking methods and allow for a continuous motion of the virtual humanoid, especially required for the tracking of avatars.

3.3.4 Motion Synthesis

The previously described methods for motion preprocessing and synthesis have been evaluated in multiple parts. First, results from the contribution to creation of footskate-clean transitions are presented. Afterwards the implications of the search strategies in the motion synthesis structure are evaluated.

3.3.4.1 Transition Generation

During preprocessing the root position is scaled with a ratio computed from the sum of the length of thigh and shank in the target and the origin skeleton. This step is necessary to avoid footskating during animation of a pre-recorded motion sequence. In order to avoid footskating during transitions only those transitions between keyframes are allowed that contain one foot planted and the other levitating. The root transformation is updated insofar as the planted foot remains in the same position. This approach is faster than spacetime constraint algorithms because it requires a computation of a transformation matrix, only. An example for an improved motion animation is shown in Fig. 3.14. The approach requires that candidates for a transition are eventually interchanged for more suitable keyframes in the neighborhood, which means that not each pair of poses can be connected with

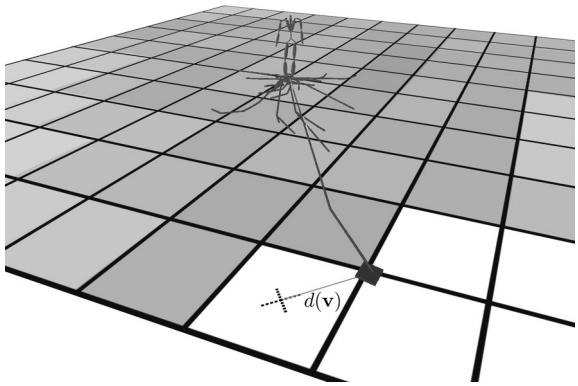


Figure 3.15. Evaluation method for the navigation capabilities of the motion graph. The humanoid is placed in the center of a plane. The motion graph is used for navigation and the final error $d(\mathbf{v})$ is measured.

a footskate-clean transition. The identification of suitable postures in the neighborhood of the candidates was, however, feasible for the majority of motion data.

3.3.4.2 Navigation

In order to evaluate the strategies for creation of motion graphs proposed two sets of pre-recorded motions are used. The first data base Π_1 consists of three walking motions with a total length of 35.8 s. In addition, the second data base Π_2 contains various other motions, like yoga, dancing, or music playing sequences, altogether 15 motion sequences with a total time of 299.7 s.

The distance error for navigation tasks is measured on a 20×20 plane without obstacles (*cf.* Fig. 3.15). Obstacles are not handled because they can be avoided by applying a path finding algorithm that creates landmarks lying on a path without obstacles. The motion graph is then used to navigate from one landmark to the next. Beginning with a virtual humanoid standing at the origin the motion graph is used to navigate to a certain position $\mathbf{v}$. To evaluate the coverage of the motion graph the reached global position $\mathbf{v}'$ is measured. For the computed motion sequence S that navigates to the desired location $\mathbf{v}$, the error is defined as $d(\mathbf{v}) = \Xi'_\Delta(S) = |\mathbf{v}' - \mathbf{v}|$ and visualized as a function of the requested goal position $\mathbf{v}$. Fig. 3.16 presents the distance error for different strategies that have been applied to the data base Π_1. Each strategy was adapted to compute a motion graph with 100 edges.

In a second step, the average distance error is compared for motion graphs with increasing size. Several motion graphs have been constructed with S_0, S_R, and S_D, all combined with the cycle strategy S_C with a varying number of iterations and used for the navigation task in a plane. The average distance error has been estimated and computed as a function of the motion graph's size. The results for the data bases Π_1 and Π_2 are presented in Fig. 3.17.

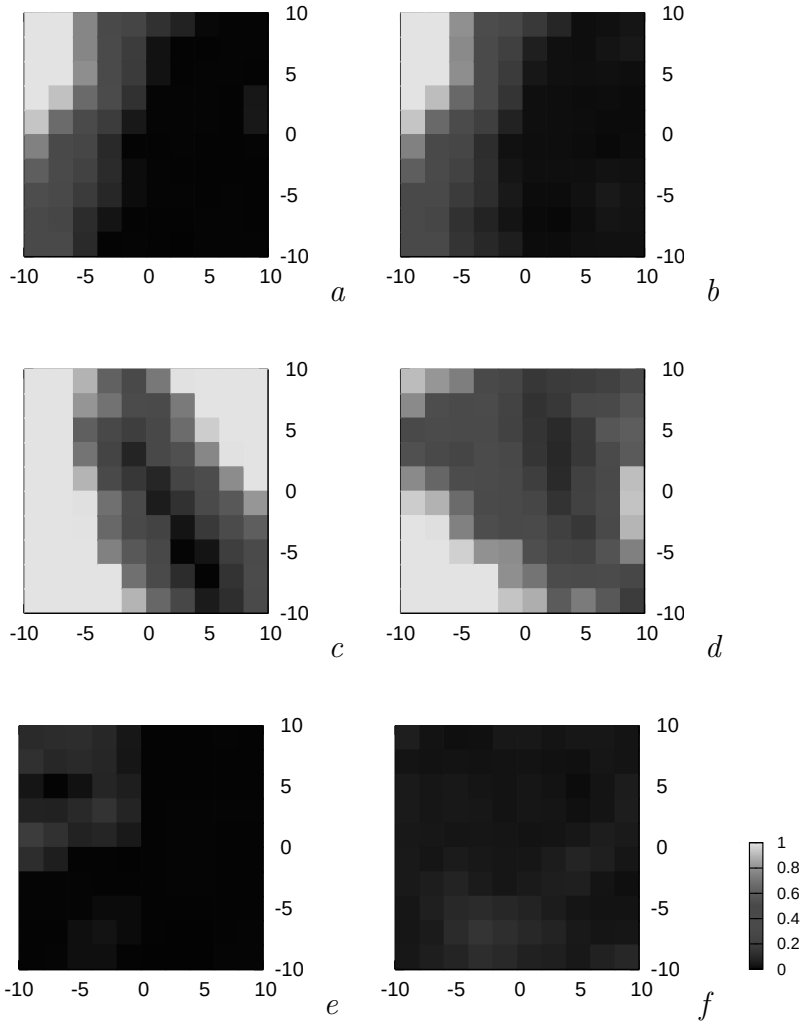


Figure 3.16. Distance error for equal-sized motion graphs created with the strategies (a) S_0, (b) $S_0 + S_C$, (c) S_R, (d) $S_R + S_C$, (e) S_D, and (f) $S_D + S_C$.

The search on the motion graph for a particular navigation query is typically processed below 100 ms. In contrast, the time for construction of the motion graphs is longer. Average values for motion graphs with a suitable distance metric of $d(\mathbf{v}) \approx 0.1$ constructed with the strategies proposed are given in Table 3.4.

3.3.4.3 Discussion

Two aspects of the motion graph-based approach to motion animation have been evaluated: optimizations to transition generation, and strategies for the construction of motion graphs.

The preprocessing methods allow for the creation of footskating-free transitions without the use of additional constraints and inverse kinematics. The transition generation approach is bound to certain keyframes with one planted and one mov-

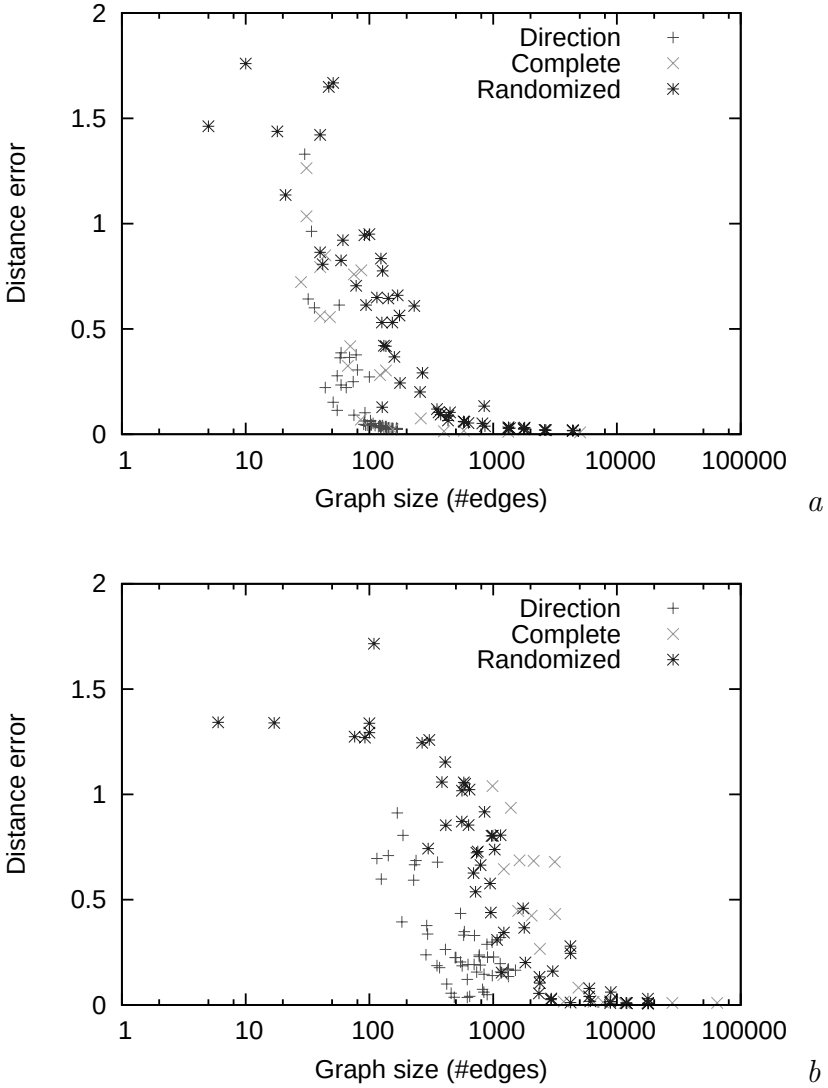


Figure 3.17. Distance error for motion graphs with varying size based on the data base (a) Π_1 and (b) Π_2.

ing foot in both sequences. However, this restriction is negligible because suitable keyframes can be found in the neighborhood of the desired transition candidate for the majority of cases. One advantage of this approach is that no additional inverse kinematics computations as proposed in [Gle01a, KSG02] are necessary in order to generate a stable position of the feet on ground. The proposed adaptation of the root position is computed efficiently and yields stable results with continuous motion.

The results from the construction of the motion graph clearly indicate that the navigation capabilities are strongly dependent on the strategy chosen during the construction process. Fig. 3.16 allows for an immediate analysis and identification of positions that can—or can not—be reached with the particular motion graph.

Table 3.4. Average construction time t and number of edges E for motion graphs with a distance metric of $d(\mathbf{v}) \approx 0.1$ created with the strategies proposed. The metrics have been calculated for the data bases Π_1 and Π_2.

Strategy	E_{Π_1}	t_{Π_1} / s	E_{Π_2}	t_{Π_2} / s
$S_0 + S_C$	869	0.7	3,658	9.8
$S_R + S_C$	166	0.3	5,789	8.2
$S_D + S_C$	60	0.2	552	2.2

As expected, the random strategy S_R and the complete analysis strategy S_0 are comparable. It can be observed, that S_0 reveals some clustering behavior because best transitions often group in one particular area. The semantic distance strategy S_D achieves the best and almost complete navigation coverage, i. e., all points in the plane are reached with a small distance error. The cycle strategy S_C is applied in addition to S_R and S_D and has an interesting effect: while the average distance error is slightly increased the overall coverage of the plane is improved. S_C can be, therefore, applied to increase the navigation coverage of the motion graph, however, at the cost of less precision. The advantage of S_D is further demonstrated in Fig. 3.17 for motion graphs of varying size. While S_0 and S_R mainly show similar behavior (S_0 showing better tendency with Π_1 and S_R performing slightly better with Π_2), the strategy S_D is generating motion graphs with better average distance error and smaller graphs. Processing time is an important issue in virtual environments. As presented in Table 3.4, strategy S_D allows for fastest construction time for a given distance metric. Even though S_D requires an additional analysis of the motion data base it is faster than S_0 or S_R mainly because it constructs graphs of a smaller size to reach the desired navigation quality.

In summary, the results clearly show that strategy S_D performs best for a fast construction of motion graphs for navigation. Good results are already achieved with smaller motion graphs in comparison to other strategies. The navigation coverage of the motion graph can be improved by application of the cycle strategy S_C, however, at the cost of an increased distance error.

The presented pipeline for synthesis of motion for virtual humanoids in virtual environments is based on motion graphs and has been improved in two stages: generation of transitions and strategies for construction of the graph. The method proposed is capable of generating motion data on the fly for a large variety of tasks. The quality of the synthesized motion has been evaluated with regards to navigation capabilities. In particular, the application of the cycle strategy and the automatic classification of direction yield improved results in comparison to the deterministic, complete analysis commonly used. To conclude, the method proposed is suitable for application in virtual environments, is fully automatic, yields fast results and creates motion sequences with a high degree of realism.

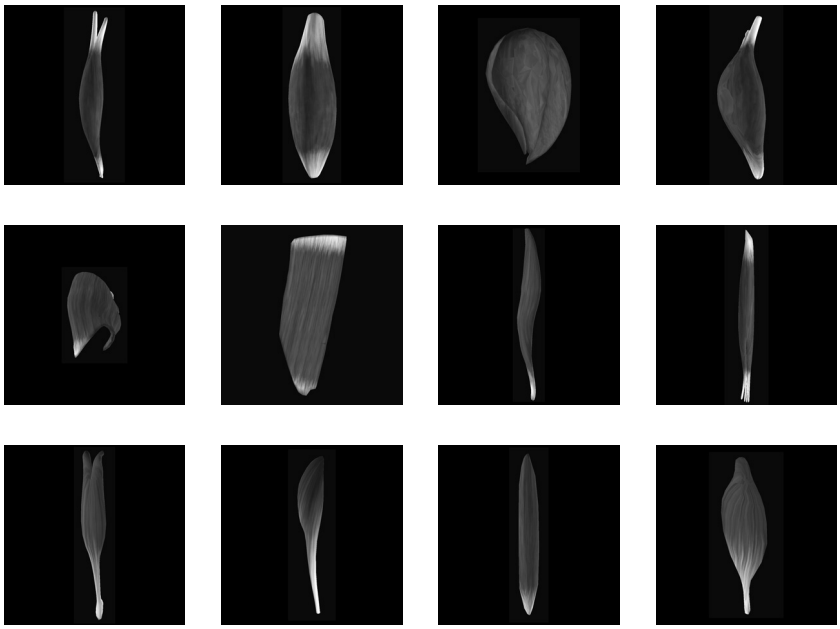


Figure 3.18. Examples of musculature geometries created on the basis of visual anatomical references. From the top left corner, line by line: the arm muscles *m. biceps brachii*, *m. brachialis*, *m. deltoideus*, and *m. triceps brachii*. The abdominal muscles *m. transversus abdominis*, and *m. rectus abdominis*, as well as muscles of the lower limb *m. extensor hallucis longus*, *m. flexor digitorum superficialis*, *m. gastrocnemius*, *m. glutaeus maximus*, *m. vastus lateralis*, and *m. tibialis anterior*.

3.3.5 Results of the Musculature Simulation and Deformation Approach

This section presents the muscle models used for an analysis of the computational cost of the musculature simulation algorithm proposed. The visual appearance and performance are compared with two frequently used techniques.

3.3.5.1 Musculature Model and Representation

To compare existing methods with the proposed deformation approach a musculature and a skeleton layer have been instantiated. Two abstract muscle data sets—mainly of human extremities—including attachment points on skeleton sites and action lines have been created. The reference data set of higher complexity consists of 30 fusiform and 22 fan-shaped muscles. The second, more simplified configuration contains 18 fusiform and six fan-shaped abstract muscles.

For an ellipsoid-based representation each of the 22 (or six for the simplified model) fan-shaped muscles have been represented by eight equally distributed ellipsoids and the remaining fusiform muscles visualized by one ellipsoid, respectively. A total of 206 ellipsoids have been used in the comprehensive model, compared to 66

Figure 3.19. The reference musculature data and geometries (*cf.* Fig. 3.18) configured on a skeleton in dresspose.

ellipsoids in the simplified model. The overall complexity of the muscle geometry adds up to 16,892 for the complex, and 5,412 vertices for the simplified geometry.

The method presented in section 3.2.3 allows for the use of arbitrary geometries. Representations based on anatomical reference have been modeled for each of the 52 abstract muscles contained in the musculature layer. A multimedia-atlas of anatomy [Pri04] has been used as a visual reference based upon the popular visible human project [SASW96] and electromyography recordings of individual patients [HBB04]. The musculature geometries were modeled in 3ds Max (Autodesk, San Rafael, CA) and assigned to the abstract musculature definition. Fig. 3.18 presents examples of the created muscle geometries that have been associated with the abstract muscle representation (*cf.* Fig. 3.19) employing two sets of geometries with varying complexity for evaluation: a basic musculature model consisting of 18,720 vertices, and a refined musculature model with 37,440 vertices.

3.3.5.2 Performance

The proposed muscle deformation and visualization approach for arbitrary geometries has been compared with established interactive techniques based on textured ellipsoids proposed by [SPCM97] and the action line visualization, both described in the Introduction. The performance of all approaches has been evaluated on a desktop-VR system, equipped with a 1.8 GHz AMD Athlon64 processor, a GeForce 6600 GT graphics adapter and running Windows XP.

The models have been tested in three different workload scenarios because the muscle geometries have been updated only if the underlying skeletal structure changed. In the worst-case scenario every muscle was deformed in each simula-

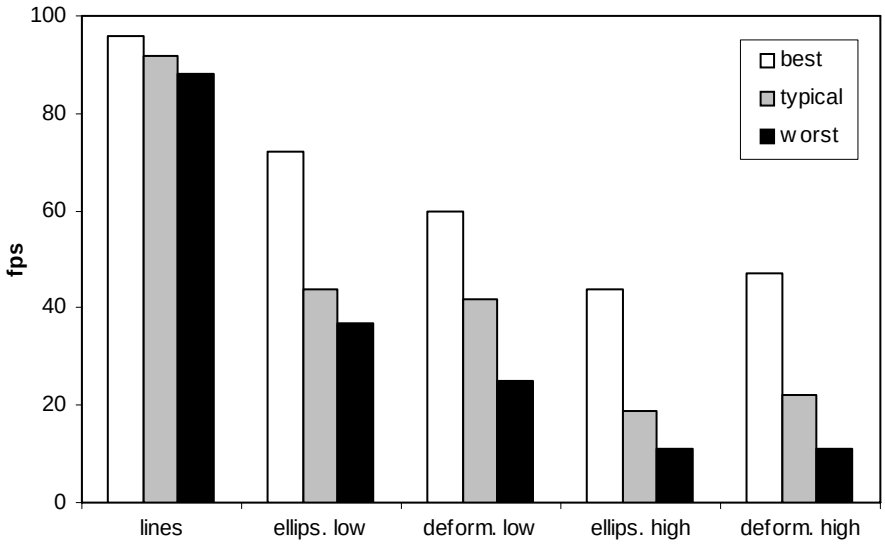


Figure 3.20. A comparison of frame rates achieved by interactive musculature visualization techniques: line and ellipsoid representation are compared with the proposed deformation approach. Ellipsoids are generated with varying geometry complexity (low and high). The deformation method is used with two musculature geometry sets of varying complexity.

tion cycle. The typical scenario was utilizing approximately 50 % of all muscles whilst no muscle was altered in the best-case scenario. The best-case scenario can thus be used as a base line reflecting the time needed to render the whole scene and to control the virtual environment. Results are shown in Fig. 3.20.

The algorithm has also been tested in a CAVE-like environment with five stereo projection walls and optical tracking (*cf.* Fig. 3.23). Within this set-up manipulations of the musculature and the underlying skeleton were possible with interactive frame rates.

3.3.5.3 Appearance

A major criterion in simulation of musculature is the final visual appearance. The visual results of the three interactive methods discussed have been compared. Fig. 3.21 shows the reference musculature model in varying poses. The muscle geometries have been deformed with the volume-preserving algorithm proposed. More detailed examples of the flexion of the *m. biceps brachii* are presented in Fig. 3.22 *a* as action lines, in Fig. 3.22 *b* as ellipsoids, and in Fig. 3.22 *c* applying the deformation method proposed.

Whilst the action line visualization facilitates a good overview, a better spatial information from a biomechanical point of view is given by the ellipsoid technique. Both visualization techniques contain an intrinsic artificial characteristic and do not

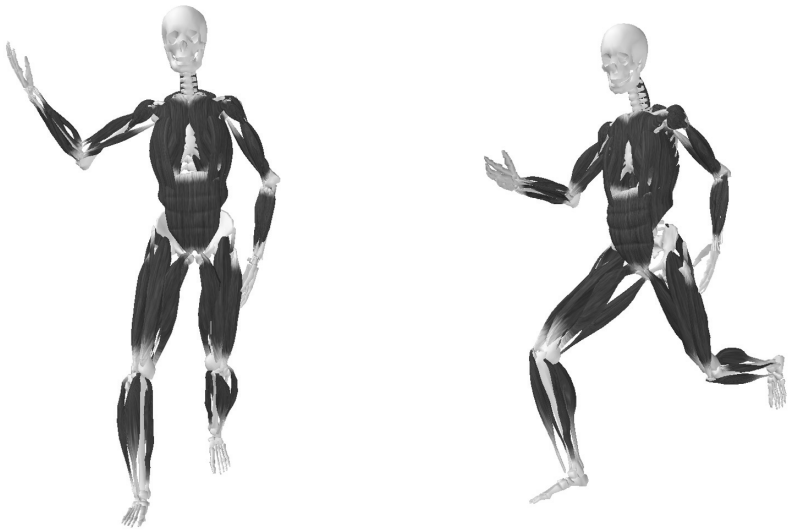


Figure 3.21. Interactive simulation of the reference musculature data (*cf.* Fig. 3.19 and 3.18) in varying poses. The musculature geometries were deformed with the volume-preserving algorithm proposed.

resemble a realistic muscle. The natural shape of musculature is visualized best with the deformation approach proposed.

3.3.5.4 Discussion

The data presented show that the technique proposed is a fast, interactive approach capable of visualizing muscle geometries with a high visual plausibility.

As indicated by the performance measurements, the method proposed is computationally comparable to methods based on geometric primitives. The most complex computations in our algorithm are in the update-routine of the volume-preserving scaling factors. As described in section 3.2.3.2, the scaling factors are computed from the changing length of the bounding box. As this computation, in fact, is only dependent on the amount of muscles and not affected by mesh-resolution, the geometry can be refined without affecting the deformation time. Topological details may be added depending on the performance of the given graphics hardware. Compared to primitive-based methods, which have to represent complex muscles with a set of geometric primitives, our method has a distinct additional advantage: it requires a single specifically shaped geometry, only.

It is obvious that by using more life-like geometries instead of primitives, the overall appearance of the visualized musculature will improve. Apart from this fact the muscle deformation has been visually compared with the pre-rendered data source mentioned in section 3.3.5.1, thereby verifying that the proposed muscle deformation is able to produce plausible results with increased level of realism of visual appearance. The technique can, therefore, be used in applications where

processing time is of concern and where simplified visualizations of geometry have been used so far. Accordingly, the technique is apt to greatly improve the level of realism without affecting interactivity.

3.4 Are the Results Obtained Conclusive in Terms of Flexibility and Interactivity?

As put forward in section 1.2, current software for virtual humanoids is often optimized for specific applications if the source code is available for adaptations at all. It has been argued that customization to varying requirements and tasks, like in medical applications, is difficult to achieve. In addition, it has been pointed out that simulation of human processes often needs time intensive computations and, therefore, requires a processing framework that offers possibilities, e. g., distribution or parallel processing, to maintain the overall system at interactive frame rates. Another consequence of the analysis of the prevailing state-of-the-art in current simulation methods for virtual humanoids led to an identification of several areas that showed potential for improvements. First, it has been concluded that methods for synchronization of audio-visual speech rendered with immersive methods, like binaural synthesis and stereoscopic visualization, principally do not exist. Second, methods for synthesis of motion graphs needed in motion synthesis techniques bear potential for further development because the prevailing automatic methods do not take the navigation quality of the resulting motion into account. Next, robustness of tracking algorithms is still an issue that has not been handled well. Elaborated methods exist for off-line processing, however, real-time systems often suffer from discontinuities and drop-outs of the tracking system. Last, simulation of musculature has been analyzed. It has been shown that most approaches are not applicable because they require time-consuming processing. Then again, interactive approaches are often focused on selected body parts or bound to simplified geometries. A flexible and interactive deformation of the overall human musculature is still missing.

In summary, the motivation for the development of a toolkit were the aspects of flexibility towards extensions and a comprehensive description for the human body, as well as interactivity, which is the basic requirement in virtual environments. In addition, several simulation methods had to be developed and integrated within the toolkit. The goal was not only a tool allowing for further neuropsychological evaluations but also a general software approach that integrates the various simulation methods for the human organism in a flexible way.

These goals were reached in so far as an improved multi-layered data structure has been developed that has proven to be flexible enough to hold a variety of data for the human body. A processing framework has been established on top of this data structure that allowed for a seamless integration of various simulation methods. This integration and the range of applications clearly demonstrate the flexibility of the software architecture proposed in this work. The framework also allows for distribu-

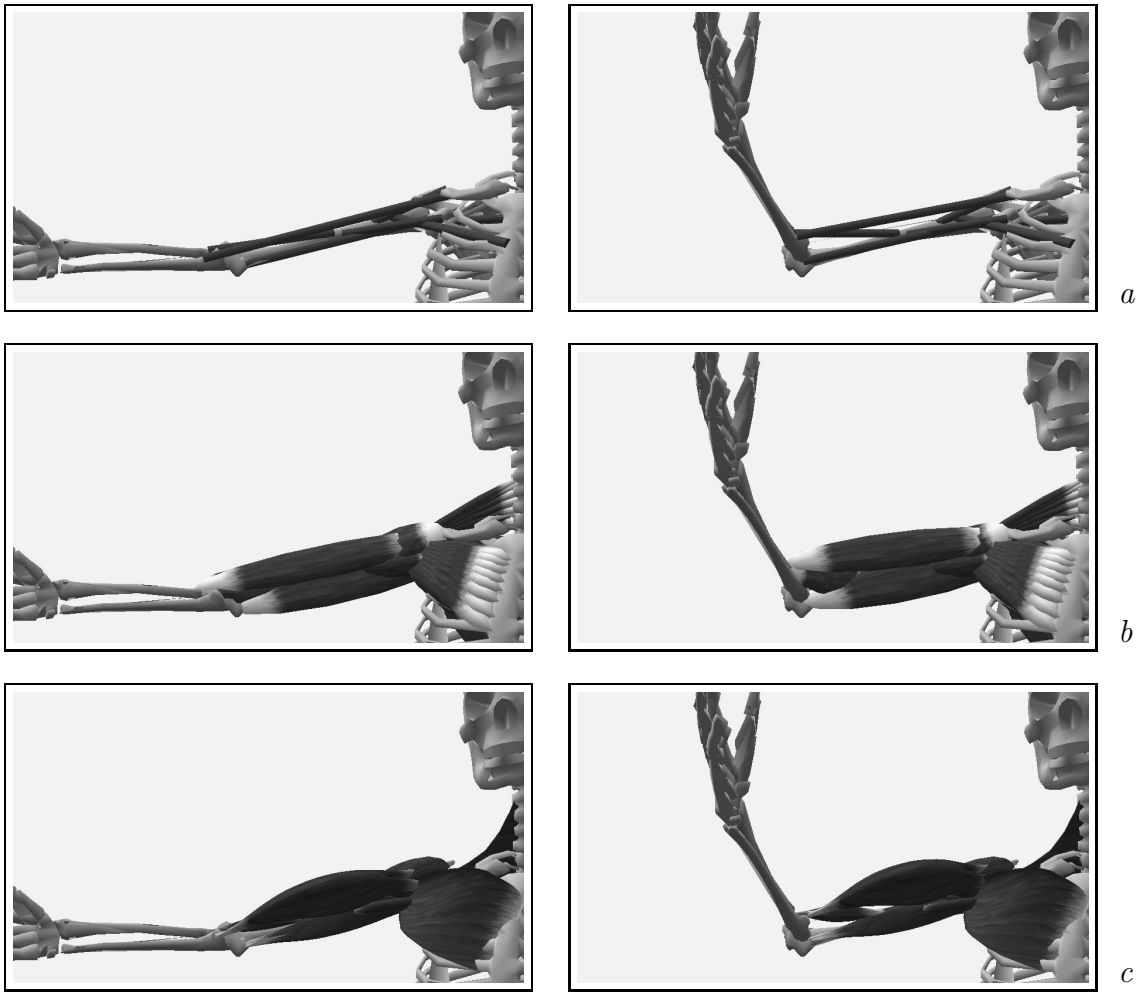


Figure 3.22. Methods for musculature visualization: (a) muscles visualized as action lines with narrow cylinders. To get an overview of overlapping musculature, this technique is the preferred choice and also provides the fastest rendering. (b) Using ellipsoids as muscle primitives as proposed by [SPCM97] yields good spatial information and illustrates the deformation in a clear way. (c) The bounding box scaling technique in VRZula gives the most natural appearance, due to the fact that arbitrary geometry is supported.

tion of the virtual humanoid and, therefore, instantiation of simulation algorithms on remote sites. The simulation itself does not affect the representation of the local copy and only the results are transferred back. It is obvious that a flexible toolkit itself can not guarantee fast computations for time intensive algorithms, however, the approach proposed here enables a flexible distribution of simulation algorithms and, therefore, allows for balancing load on computational resources.

The previously mentioned open issues in methodology have been solved by specific novel contributions or improvements to established techniques. The complex deformation of musculature has been reduced to the fast computation of bounding

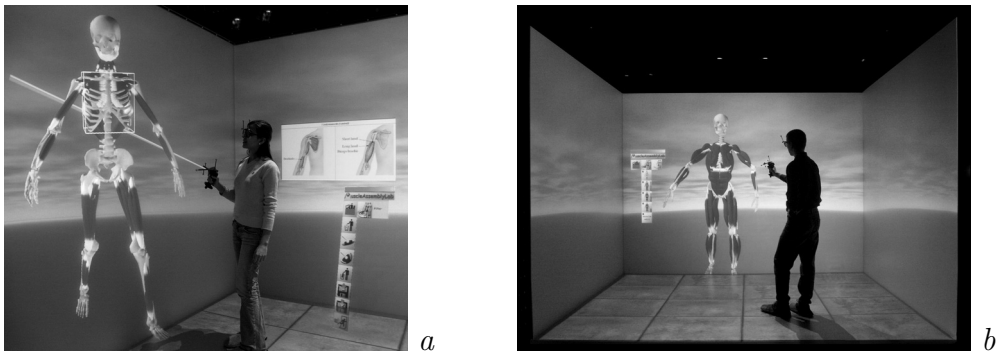


Figure 3.23. Representation and manipulation of musculature in a CAVE-like environment: muscles are represented as (*a*) ellipsoids, and as (*b*) anatomy-based geometries enabled by the algorithm proposed.

boxes. It has been shown that this approach preserves the volume of arbitrary musculature geometries. It greatly improves the degree of realism in appearance while showing the same performance as established interactive techniques. Audio-visual speech representation has been realized as a distributed system. The synchronization is based on a performance analysis that was further used for identification of bottlenecks in the synthesis process. Altogether the modular system enables interactive synthesis of multimodal and immersive speech. The motion tracking approach proposed handles discontinuities of tracking data by a threshold analysis on two levels. Drop-outs of sensors do not lead to jumps in the kinematics of the tracked virtual humanoid but, in the worst-case, results in a smooth delay perceived as latency of the tracking system. Several strategies for the synthesis of motion based on motion graphs have been proposed and evaluated. It has been shown that a fast automatic analysis of the motion data yields best results for navigation, i. e., smaller and faster motion graphs are required for optimal results. The algorithms proposed have been integrated in the VRZula framework and represent real-time capable solutions for the particular tasks that facilitate the use of virtual humanoids in virtual environments.

To conclude, the goal of a comprehensive simulation of the human organism is still beyond reach and far too complex in all its details for current computational resources. However, the approach proposed in this work delivers powerful methods for a flexible integration and interactive realization of partial solutions in that great puzzle.

4

Influence of Avatars in Virtual Environments

The previous chapter has described methods that were researched and developed in the thesis and that were integrated in the flexible, interactive toolkit for virtual humanoids, which has enabled several experimental evaluations regarding the application of virtual humanoids as avatars in virtual environments. As clearly evident from the foregoing technical discussion on the use of interactive virtual humanoids in virtual environments the following question remains: which effects has the use on the mental processes of a user during perception of and interaction with the virtual environment? As discussed and motivated in section 1.3 and chapter 2, the application of virtual humanoids as avatars holds considerable promise and seems to be capable of positively influencing the sensation of presence in users. It has been proposed that a high sensation of presence leads to an amelioration of subjective user preference toward a particular medium and thus has a positive influence on usability. The working hypotheses focusing on application of avatars and their influence on presence in users are presented in the sequel together with two studies that contribute to the clarification of these interrelations.

4.1 Working Hypotheses

The sensation of presence is influenced by a variety of factors. While research on technical issues of virtual environments has not led to a satisfactory increase of presence and even seems to be trapped in the "uncanny valley" (*cf.* section 1.3), the attention has been turned towards the content of virtual worlds. The interaction

with representations of humans in virtual environments is assumed to ameliorate presence, in particular, self-presence and social presence. It has been discussed that virtual humanoids are capable of increasing presence, e. g., in [Bio97], yet this presumption is mainly based on theory, since experiments have not satisfactorily clarified this interrelationship. Thus, the application of virtual humanoids as avatars is promising but lacks thorough experimental proof.

The influence of avatars on presence is determined by factors involved in the process of interaction between a user and his avatar. The perception of an avatar and the inherent proprioceptive feedback create a sense of identification with the avatar of lesser or larger quality. Data from external senses are compared to real world stimuli and to the current configuration of the own, real body that is provided by the internal sense of proprioception. Results from this processing are integrated in the mental world representation and can lead to an identification of the avatar with the representation of the own body, the body schema. As of now, the influence of the mapping of kinematics of the user's body on the avatar on mental processes in users is not well understood.

Another important issue in the evaluation of the influence of avatars on users is imposed by the virtual environment technology itself. It seems that an avatar is preferably accepted as a representation of the own body if located in the same spatial position as the own body, and if perceived from the natural egocentric view, i. e., from the first person perspective. Such a scenario can be generated with head-mounted displays (HMD) where the vision of the real world is completely replaced by virtual stimuli. However, current virtual environments are room-mounted installations where a user perceives himself surrounded by the virtual world. In such a scenario a matching of the avatar to the location of the user makes no sense because it creates a perturbing interference between the real and virtual stimuli. Hence, an exocentric, i. e., a third person perspective has to be used, instead. David and colleagues have reported results indicating that variations of perspective lead to a decreasing quality of agency, a sense to be in charge of action [DBC+06]. By contrast, Taylor points out that third person perspective in 3D games can contribute to the involvement in the mediated environment because the avatar seen from a third person perspective offers an identification possibility within the context of the virtual world [Tay02]. Given the fact that most studies dealing with avatars and presence have been carried out in egocentric scenarios, and given the conditions of current room-mounted virtual environments requiring an exocentric view on avatars, it is of interest to further examine the capabilities of avatars in room-mounted virtual environments.

The last open issue deals with presence and usability. As discussed in section 1.3.2, Bowman *et al.* clearly see presence as a user preference that generally contributes to usability [BKLP05]—a view that has not been unequivocally substantiated and needs, therefore, further experimental validation.

Hence, the following conceptual points were considered essential working hypotheses for a deeper understanding of the application of virtual humanoids as avatars in virtual environments:

Kinematics Hypothesis. Synchronous kinematics increase subjective preference for an avatar.

Presence Hypothesis. Avatars in virtual environments are identified with the body schema and increase the sensation of presence.

Usability Hypothesis. Presence influences usability of virtual environments.

The hypotheses were tested with experiments described in sections 4.3 and 4.4. In addition, a structured approach for VR-based neuropsychological experiments is presented in the next section.

4.2 NeuroMan – Neuropsychological Experiments in Virtual Environments

The neuropsychological studies undertaken have been realized with the virtual humanoid technology presented in chapter 3. In order to create a structural basis for the experiments the NeuroMan system has been developed that combines a clear definition of the experimental set-up with the execution in virtual environments. As pointed out in [VKW+05] and [WAVK07], the majority of integrated approaches[1] is focused on experiments performed with desktop monitors or head-mounted displays. Immersive representation, i. e., stereoscopic visualization, user-centered projection, and tracking are often missing. Game engines are used as experimental environments, as well. They allow for the definition of an experimental session, however, the resulting experiment is closely bound to the pre-arranged gaming environment. To date, there is no clearly defined system that allows for a seamless integration of neuropsychological experiments in immersive, multimodal, and interactive virtual environments. This section briefly describes the solution proposed, as well as the application of the NeuroMan system in several neuropsychological studies.

4.2.1 Software-based Organization of Neuropsychological Experiments

The NeuroMan system defines an experimental structure consisting of sessions, blocks, trials, and scenes. The whole structure is determined prior to the experiment execution and is mainly defined by a script. A scene is specified as a list of multimodal stimuli, i. e., virtual objects, humanoids and sound properties. Scenes are grouped by trials that define the chronology of the representation of the scenes and their behavior upon user interaction. Modifications to objects and reactions

[1] Prominent commercial examples are Presentation (Cortech Solutions, Wilmington, NC), or E-Prime (Psychology Software Tools, Pittsburgh, PA).

of the system during runtime are defined as script functions, as well. Blocks organize the various aspects of execution of trials, like randomized or sequenced execution, number of iterations, *etc.* Finally, sessions are used to organize the whole experiment. The overall experimental structure of NeuroMan has been described in [VAD+03, VKW+04, VKW+05]. The system also allows to directly set several subject-specific data, most important being the interpupillary distance often neglected and wrongly set to mean values in other approaches [Dod04].

The NeuroMan system records all user interactions, like trajectories, or reactions to events with a high accuracy. It uses a pipes-and-filters approach to monitor and analyze input data with low latency by application of several light-weight processes [VAD+03]. This approach has been extended to detect timing instabilities of the operating system and realizes a method for a software-based detection of reaction times with a standard deviation of less than 1 ms, which is sufficient for the majority of experimental tasks [WAVK07]. For more precise measurements of reaction times the system includes a hardware-based detection mechanism that has a precision below 0.01 ms [VAKB04].

4.2.2 Application of NeuroMan

The NeuroMan system has been applied and evaluated in a number of experiments. The following list gives a brief description of the research topics in each evaluation.

Depth Perception. This study is focused on the quantification of depth perception in virtual environments. Depth is underestimated significantly in virtual environments irrespective of depth cues [AHV+05].

Mental Processing of Numbers. This study examined the processing of numbers perceived in a virtual environment. Reaction times have been measured with the hardware approach of NeuroMan. The results contributed to the understanding of the spatial alignment of sequenced numbers [GNV+04]. This research has been currently extended by an fMRI-guided transcranial magnetic stimulation study that is focused on the relevance of object features such as numerical meaning on the perception of 3D space.

Asymmetry of Attention. Attentional processes in virtual environments have been researched in this study. An attention asymmetry under sleep deprivation has been found [HVK+04, HVK+05]. Further work on covert attention realized with NeuroMan has been reported in [HWKF07].

Navigation in Virtual Environments. This study evaluated navigation and orientation processes in a HMD-based and a CAVE-like virtual environment [And06]. Preliminary results clearly show the advantages of room-mounted virtual environments.

Performance Study. The user performance in a CAVE-like environment has been measured and compared to results achieved with a desktop-VR system. Subjects had to accomplish a docking task that consisted of the creation of a musculature model on a virtual skeleton. This within-subject study revealed a significant increase in the time performance without loss of precision in the room-mounted set-up [VKW+05].

Near and Far Space. The project deals with neural correlates of spatial processing in peripersonal and extrapersonal space [BVW+07]. Spatial everyday tasks were created in a virtual environment and synchronized with fMRI measurements. The project is focused on application in diagnostics and rehabilitation of neglect symptoms.

Crossmodal Cueing. A three-dimensional cube-shaped paradigm has been developed that serves for the assessment of visual, auditory, as well as cross-modal cueing effects on spatial orientation of attention [AWKS07].

Prehension. An experimental set-up has been created in which intuitive prehension movements for virtual objects are recorded and analyzed. The project is focused on general aspects of grasping without haptic feedback in virtual environments and aims at answering the question whether grasping in virtual environments can be compared with real grasping [AWK+07a, AWK+07b].

4.2.3 Discussion

The proposed framework has been motivated by the fact that an integrated approach to neuropsychological experiments in virtual environments has not been proposed and realized so far. The system described in the last sections defines an experimental structure and allows for a highly sophisticated timing control for reaction time measurement and input event monitoring. The system is directly bound to a VR software allowing for immersive, interactive and multimodal experimental evaluations. The concepts behind the system and its input devices have proven their applicability in a number of studies and have also been successfully used in the studies described in the next sections.

4.3 Virtual Arm Study

The goal of the *virtual arm study* was to investigate cerebral activations linked to perception and processing of partially mismatched kinematics. It should clarify the extent of acceptance of a virtual humanoid representation as the own body, i. e., whether representational stimuli generated by a virtual humanoid are capable of inducing perceptual processes normally found in reality.

Problems during substitution of reality by computer-generated environments and the influence on the mental processing have been discussed in sections 1.3 and 1.4,

revealing the lack of any functional imaging study on avatar-mediated motions and correlated aspects of synchrony. Therefore, the subsequent study has been designed to explore the effects of avatars on healthy subjects. Subjects have been asked to perform simple movements that have been mapped on a virtual humanoid. The resulting trajectories, as well as activations in the brain have been observed. A well known deformation in kinematics during the task of drawing circles has been chosen as metric: normal subjects are, in general, capable of performing visually controlled motion with great accuracy. Some tasks like the previously mentioned drawing of circles, however, impose an error in the resulting trajectory. The circles are not perfectly shaped but have an elongated, ellipsoidal shape [SLT86]. Soechting *et al.* conclude that the deviation of the hand trajectory is based on the relationship between joint angles caused by non-sinusoidal torque patterns [ST86]. These observations are correlated to biomechanics of the arm by Pfann and colleagues [PCMH02]. It is concluded that biomechanics only is the cause for the elongation. Other studies, however, suggest that subjects are able to take inertial anisotropy into account, linking the whole task to perceptual processes [SJW98, FL01]. Taken this evidence together, it is still unclear if the elongated shape of circular movements is pure biomechanical or due to a perceptual effect.

This study used this discrepancy to induce conditions with optical and kinesthetic sensory mismatch in order to explore the effects of the perception of an interactive arm avatar. While subjects performed circular movements with their right arm, the virtual arm was represented either as a left or a right limb. Furthermore, the kinematics of the virtual arm was directly mapped on the virtual arm or mirrored on the mid-sagittal plane. If the elongated shape would be a pure biomechanical effect, this should not change with alteration of the visual feedback. If the virtual arm would be capable of inducing perceptual processes and, therefore, have a perceptual component, this should be reflected in kinematic changes in dependence of the visual information. In addition, the entire experiment was performed while subjects were lying in a PET scanner in order to measure their brain activity during movement execution. This allowed for further implications based on correlation of kinematic changes with cerebral structures.

4.3.1 Methods

Participants. Ten normal, right-handed subjects (as assessed with the Edinburgh handedness inventory [Old71]) with no history of medical or neurological disease (five males and five females, age: 19 to 42 years, mean: 24.4 years) participated in the study. All subjects had normal or corrected-to-normal vision and were not previously briefed about the purpose of the study. The study was approved by the local ethics committee.

Design. Subjects were always required to perform movements with their own right arm. They perceived a virtual limb whose motion was directly linked to the real arm.

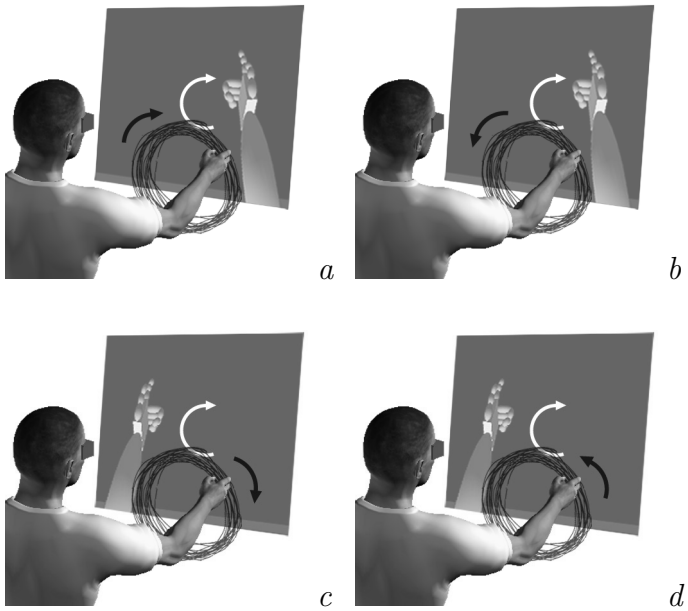


Figure 4.1. Study design: the four experimental conditions (*a*) RA–, (*b*) RA+, as well as (*c*) LA–, and (*d*) LA+ with variation in the represented arm and kinematic mapping. Subjects observed the virtual arm avatar through an HMD and, therefore, could not see the real body.

Two main factors were manipulated in this study: the laterality of the virtual limb, either a left (LA) or a right virtual arm (RA) ('Arm' manipulation); and the motion, which was either directly mapped (–) or mirrored (+) ('Mapping' manipulation) at the subjects' mid-sagittal plane. Both factors were crossed in a 2×2 within-subjects design resulting in four different experimental conditions: LA+, LA–, RA+, and RA– (*cf.* Fig. 4.1). It should be noted that this does not reflect a fully factorial design. Even as the two factors could be varied independently, they are not perceived as independent from each other. The RA– condition is a direct reproduction of motion and LA+ corresponds to those observed through a mirror; both reflect, therefore, the usual, canonical experience in reality. Consistently, subjects reported conditions RA+ and LA– as more 'awkward', as they can not be related to movements of the own body perceived in the real world. Thus, these two conditions has been designated as non-canonical.

Procedure. The experiment consisted of one introductory, one training and one experimental sessions. During the introductory session, subjects were informed about the course of the experiment and interrogated about their medical history and their handedness. After this procedure, they gave their written consent and a blood sample was drawn for same-day analysis. The interrogation took place at least one day prior to the PET experiment and provided inclusion and exclusion criterions of the subjects.

During the training session (which usually took place immediately after the introductory session), subjects were acquainted with the technical set-up: they were seated upright, provided with an HMD and a tracker device was placed on their right arm on the back of the hand. The subjects were able to freely perform arm movements in front of them. Subsequently, two learning tasks consisting of touching virtual objects with the virtual arm under the four conditions were performed. The session lasted several minutes until the subjects reported to have achieved a good 'feeling' for the virtual arm movements.

The experimental session consisted of the actual PET measurements. After an intravenous injection of radioactive markers the subjects were placed in a PET gantry with their heads fixed in a personal-fitted head rest. They were wearing an HMD and their right arm was tracked in the same way as during the training session (*cf.* Fig. 4.2). Arm movements had then to be performed above their chest level in an arm position of about 45 ° inclination. Subjects were then instructed to perform movements with their right arm in such a way that the virtual arm described a vertical clock-wise circle at a frequency of 1 Hz. Thus, subjects had actually to perform clockwise movements of their own arm in the RA– and LA– conditions and counter-clockwise movements in RA+ and LA+ conditions in order to generate the desired visual image of a clockwise-rotating virtual arm. One trial took 60 seconds and subjects performed three repetitions of all four conditions, summing up to a total number of twelve trials per subject. Trials were separated by seven to nine minutes of rest between the single trials, allowing the radioactive tracer to decay. Additionally, all subjects received a structural MR scan for stereotaxic normalization of the PET data and exclusion of potential cerebral abnormalities.

Apparatus. The regional cerebral blood flow (rCBF) was measured after injection of O^{15}-butanol as radioactive tracer. PET measurements were made with a Siemens ECAT EXACT HR+ PET scanner with a total field of view of 155 mm covering the whole brain.

The virtual environment consisted of a V8 stereoscopic HMD (Virtual Research Systems, Aptos, CA) and a Fastrack electro-magnetic tracking device (Polhemus, Colchester, VT), forming an immersive (stereoscopic), interactive environment. A Dual Pentium II (450 MHz) Intergraph PC with two Wildcat GPUs running Windows 2000 was used to generate the stereo images and to record tracking data. The environment was driven by an early version of the ViSTA software based on Sense8 WTK.

The virtual arm was modeled as an articulated structure and represented with geometric primitives (ellipsoids and boxes). The hand was modeled with its index finger outstretched and the other fingers flexed. This virtual arm could either be a left or a right arm, differing (1) by the position of the shoulder with respect to the subjects' viewpoint and (2) by the flexing direction of the fingers. The virtual arm was represented on a spatially shifted position viewed from a first person perspective (*cf.* Fig. 4.1 and 4.2). An analytical inverse kinematics approach was applied in order

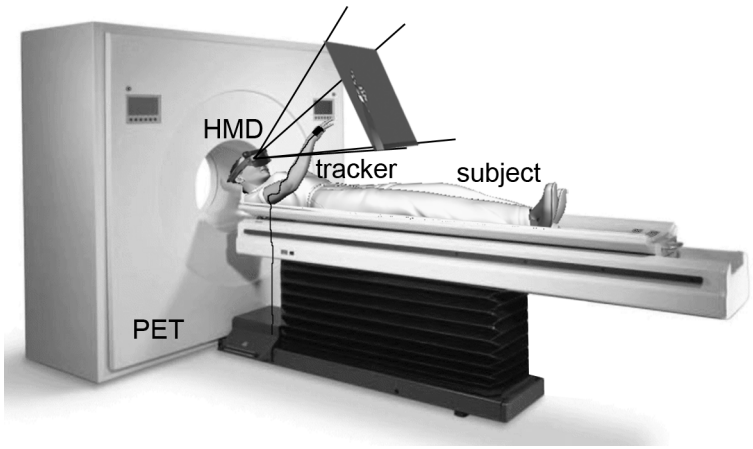


Figure 4.2. The apparatus set-up: wearing an HMD the subjects are laid in a PET scanner. A tracker device is fixed on their right arm and determines the motion of a virtual arm.

to simulate the motion of the virtual limb (*cf.* section 3.2.2.1). The end effector position was determined by the tracking device. The virtual arm was presented to the subjects by means of the HMD.

Data Processing and Analysis. For each subject, 12 scans were measured for a period of 40 s after injection (555 MBq per scan). Bolus injections were performed into the left brachial vein and immediately flushed with 10 ml saline. The whole study involved the administration of 5 mSv effective dose equivalent of radioactivity per subject. Analysis of the rCBF data was done with SPM'99. Image data were realigned to the first emission scan and co-registered to the individual subjects' MRI. The analysis was performed in the following way: first, the main factors 'Arm' and 'Mapping' were calculated. Second, the influence of the factor 'Mapping' was calculated for right arm and left arm trials separately. Only activation clusters with at least 20 voxels significant at $p < 0.0001$ (not corrected; $Z > 3.87$) were considered.

The subjects' movements were sampled at approximately 25 Hz and stored for off-line analysis. The data were analyzed using customized Matlab modules (The MathWorks, Natick, MA). For each trial, the absolute path length of a trajectory Λ, and the mean velocity v were calculated. As anticipated and discussed before, subjects did not perform exactly circular movements but created trajectories of a rather elliptical form.

To find the best approximating ellipse for a particular trajectory the data has been projected on the fronto-parallel plane and a principal component analysis was performed that yielded the main axes and the eigenvalues. In a subsequent step a direct optimization approach was used to search for an ellipse aligned with the main axes, and with a ratio ϕ adapted to the eigenvalues that had the minimal

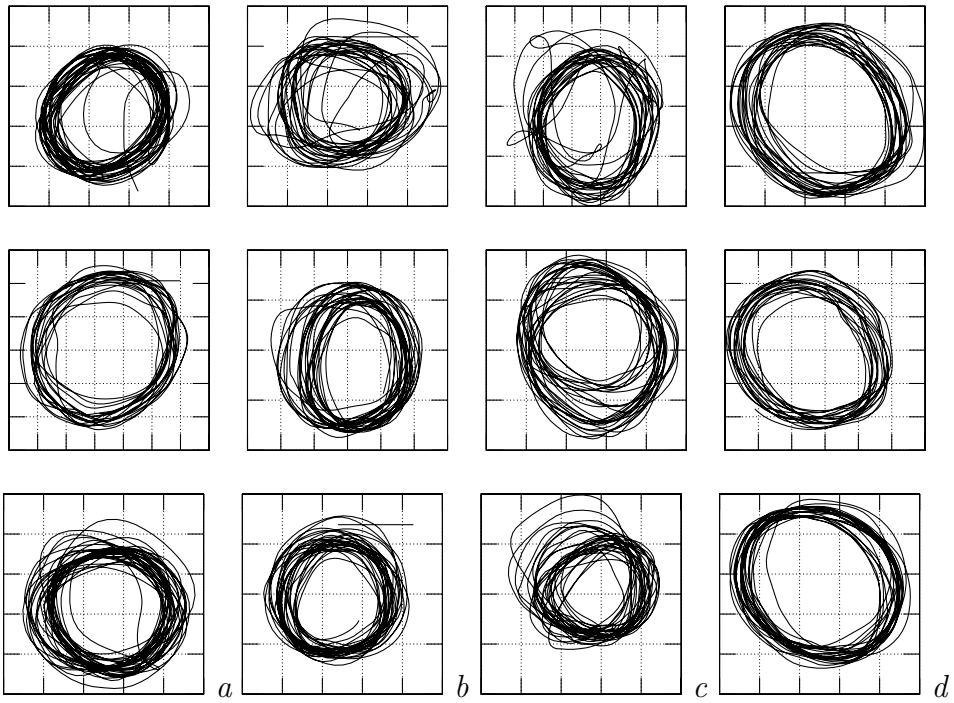


Figure 4.3. Examples of trajectories generated by subject BA02 projected on the fronto-parallel plane. One tick on each axis represents the distance of 10cm. Each column shows results from a condition: (*a*) LA+, (*b*) LA–, (*c*) RA+, and (*d*) RA–. Columns (*a*) and (*d*) show the canonical conditions that can also be found in reality. The rows show the first, second, and third repetition of each corresponding trial. All trials have been carried out in a pseudo-randomized order and resorted in this figure.

distance error ρ to the trajectory. ϕ was chosen as the ratio between the first and the second eigenvalue, therefore, $\phi \geq 1$. ρ was computed as the mean of the distance between each sample point on the trajectory and the closest point on the ellipse. In the case of nearly circular trajectories, i. e., $\phi \approx 1$, the calculation of the tilt is not meaningful. Therefore, the inclination of the ellipse γ was only submitted to statistical analysis for $\phi > 1.15$. Because the main axis has no direction γ was normalized to the range between $-90°$ and $90°$. In summary, the fitting analysis provided the parameters ρ, denoting the accuracy of repetition, the ratio between the main axes ϕ, and for trajectories with $\rho > 1.15$ the inclination of the main axis of the ellipse γ. These parameters were supplied to a two-way ANOVA with the factors 'Arm' and 'Mapping'.

4.3.2 Results

Subjective Reports. During the learning session subjects were asked to describe their experience with regard to the virtual arm and the kinematics. No subject had difficulties in spontaneously identifying the virtual arm as a model of a right or left

Table 4.1. Mean values and standard deviations of trajectory kinematics for all conditions. Λ denotes the path length, v the mean velocity of the trajectory. Both measures are given in millimeters. No significant differences have been found between the conditions.

	Condition			
Measure	LA+	LA–	RA+	RA–
Λ / mm	31,340 ($\pm$ 14,210)	33,250 ($\pm$ 15,520)	31,180 ($\pm$ 12,970)	30,490 ($\pm$ 12,720)
v / $^{\text{mm}}/_{\text{s}}$	346 ($\pm$ 128)	348 ($\pm$ 128)	335 ($\pm$ 105)	334 ($\pm$ 121)

human arm, nor did any of the subjects report the experience of a delay between their own real movements and the movements of the virtual arm.

Kinematic Data. The form and kinematic parameters of the motion trajectories (*cf.* Fig. 4.3) have been analyzed. For technical reasons, though, only 103 out of the 120 recordings could be used for the kinematic analysis.

The average values of the kinematic parameters for all conditions are shown in Table 4.1. The two-way ANOVA revealed no significant effect of any factor on the path length Λ (factor 'Arm': $F(1,9) = 0.48$, n. s., factor 'Mapping': $F(1,9) = 0.68$, n. s.) and mean velocity v (factor 'Arm': $F(1,9) = 0.01$, n. s., factor 'Mapping': $F(1,9) = 0.82$, n. s.). Both measures indicate the actual motor performance to be similar over all conditions.

As mentioned before, the inclination of the ellipse approximation was only considered for trials with a ratio $\phi > 1.15$, which was the case in 87 out of the 103 valid trials. The average kinematic parameters across all subjects are shown in Fig. 4.4 and detailed in Table 4.2. Interestingly, the fitted ellipses of the trials with a virtual left arm were tilted more to the left than the trials with a virtual right arm. Furthermore, the fitted ellipses of those trials with movement mirroring were more elongated and less accurate than the trials without mirroring, irrespective of the laterality of the presented virtual arm. This was reflected in the statistical analysis: the factor 'Arm' had a highly significant effect on the inclination angle γ ($F(1,9) = 16.05$, $p < 0.001$), but not on accuracy ϕ ($F(1,9) = 0.618$, n. s.) or on ratio ρ ($F(1,9) = 1.20$, n. s.). Contrary, the factor 'Mapping' proved to have a significant effect on accuracy ϕ ($F(1,9) = 5.64$, $p = 0.02$) and ratio ρ ($F(1,9) = 33.35$, $p < 0.001$), but not on the angle γ ($F(1,9) = 0.09$, n. s.). For all comparisons, the interactions were not significant.

Cerebral Activations. The results of the SPM'99 analysis are shown in Fig. 4.5 and Table 4.3. The comparison of the conditions LA *vs.* RA and RA *vs.* LA, respectively, with collapsed 'Mapping' manipulation showed one single, extended activation in the precuneus, i. e., in the superior occipital cortex, at the border to the

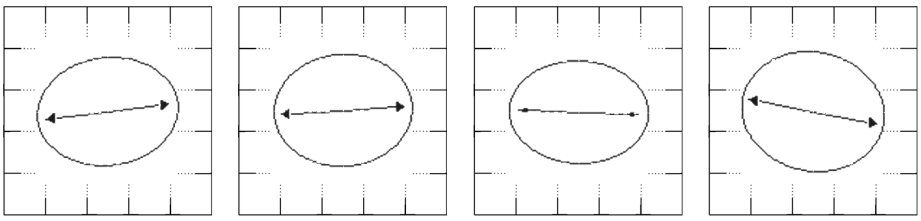


Figure 4.4. Principal components and the final ellipsoid approximation of the kinematic trajectories (*cf.* Fig. 4.3 for an example) in the four conditions (from left to right) LA+, LA–, RA+, and RA–. The tilt of the main axis is dependent on the laterality of the represented virtual arm.

posterior parietal cortex. This activation was symmetrical and strictly contralateral to the side of the virtual arm. Fig. 4.5 shows the effects for the activation clusters. For both precuneal regions, the activation strength does not follow a fully factorial scheme, but is modulated in the order RA– > RA+ > LA– > LA+ (and *vice versa*).

4.3.3 Discussion

Taken the results of the kinematic and the imaging analysis together, it is evident that the form of circular movements is a result mainly caused by perceptual processes mediated by lateralized precuneal activations. These processes have been induced by a virtual arm avatar identified with the body schema.

4.3.3.1 The laterality of the virtual arm affects parameters of kinematics

The kinematic results show the orientation of the main axis to be dependent on visual feedback: based on the laterality of the visualized arm the main axis is tilted to the left or to the right side, respectively (*cf.* the two left *vs.* the two right columns in Fig. 4.4). Interestingly, the effect is solely dependent on the laterality of the represented arm and is not changed through the mirroring on the mid-sagittal plane. This allows the conclusion that the orientation of the ellipse is to a major extent dependent on visual perceptual processes rather than biomechanical constraints.

This seems to contradict the results of the study of Pfann *et al.*, who concluded that performing circles in a more oval form is due to the lack of time for the CNS to compensate for biomechanical anisotropies [PCMH02]. Pfann's paradigm, however, differs from our one in one important point: her subjects were instructed to concentrate on the path of a black dot on a transparent pointer attached to the forearm, whereas ours focused on the movements of the visually presented limb. Furthermore, Pfann and colleagues have focused on the speed of the motion task and draw their conclusion based on the variations towards the fastest performance. For motion with slow pace, Pfann's results agree with the influence of perceptual processes. In our

Table 4.2. Parameters (mean value and standard deviation) of the ellipsoid trajectory approximation for all conditions. ρ denotes the accuracy, i. e., the distance of the optimal ellipse to the trajectory, ϕ the ratio between the two main axes, and γ the tilt of the main axis. The tilt has been analyzed for ellipses with $\phi > 1.15$, only.

	Condition			
Measure	LA+	LA–	RA+	RA–
ρ / mm	15.8 ($\pm$ 2.4)	14.0 ($\pm$ 2.9)	15.4 ($\pm$ 3.2)	14.3 ($\pm$ 3.1)
ϕ	1.34 ($\pm$ 0.21)	1.25 ($\pm$ 0.16)	1.36 ($\pm$ 0.16)	1.20 ($\pm$ 0.1)
γ / deg	10.6 ($\pm$ 17.5)	1.3 ($\pm$ 22.5)	-3.0 ($\pm$ 17.3)	-6.2 ($\pm$ 36.6)

Table 4.3. Centers of activation foci for the main factors. * denotes a significance at a level of $p < 0.05$ (corrected; $Z > 4.9$).

		MNI coordinates				
Main Factor	Region	x	y	z	Z	p
LA *vs.* RA	left *precuneus*	8	-76	46	6.88	*
RA *vs.* LA	right *precuneus*	-10	-82	40	6.45	*

study, subjects have been asked to perform circles at 1 Hz, which is slightly faster than the 0.7 Hz, the average of the slowest pace in Pfann's experiment.

In our study, the values for elongation and orientation in the most natural condition RA– are comparable to those obtained in other studies, particularly with respect to the dependence of the inclination of the ellipsoid on the laterality of the observed arm. Our results seem to confirm Soechting's original proposal that "the topology of the sensorimotor map used for the production of the movement and for its perception is the same" [SLT86]. Furthermore, this finding provides additional evidence for the hypothesis that movements are mainly guided by perceptual cues rather than by internal feedback mechanisms [MKKP01].

It is safe to conclude that the CNS tries to perform circular movements with optimized energetic effort. The form of these optimized trajectories is regarded as natural and already implemented at the perceptual level. Our subjects performed their movements in such a way that the movements of the virtual arm were looking as natural as possible. As these perceptual processes were induced by a virtual arm, it may safely be concluded that an avatar with synchronous kinematics is accepted as the own body, i. e., is identified with the body schema.

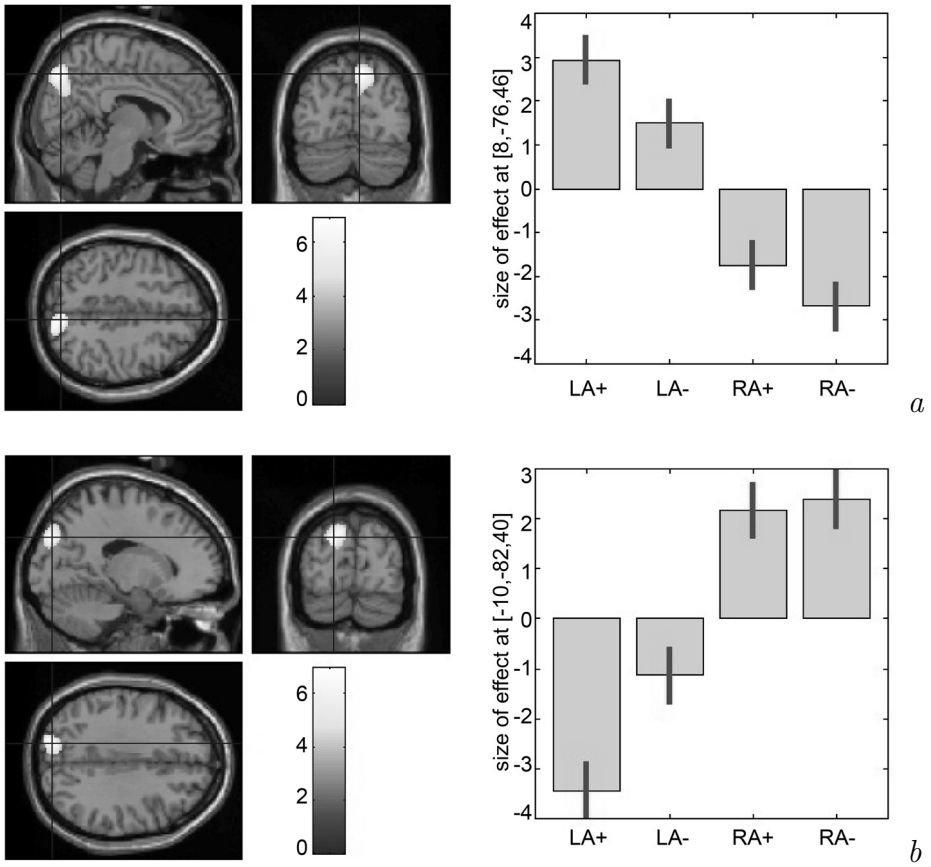


Figure 4.5. Activation clusters and corresponding parameter estimates in the *virtual arm study*: cerebral activation is superimposed on sagittal, coronal, and axial sections of the reference MR data set. Comparison of conditions (*a*) LA *vs.* RA, and (*b*) RA *vs.* LA with collapsed 'Mapping' manipulation yielded one main cluster in the precuneus area strictly contralateral to the represented virtual arm (*cf.* Table 4.3).

4.3.3.2 The virtual arm activates the precuneus that integrates visual and kinesthetic information

The influence of an avatar on visuo-kinesthetic processes and the integration in the body schema can be, furthermore, substantiated by evidence from the analysis of cerebral activations that makes it possible to correlate the previously discussed changes in kinematics to specific cerebral structures.

The results clearly show that differences in parameters of kinematics are accompanied by strictly lateralized activations in the precuneal region in both hemispheres. This corresponds to findings by Dohle and colleagues who found lateralized activations in a similar set-up: subjects were asked to observe their own (real) hand either directly, or through a mirror, resulting in conditions corresponding to RA+ and LA– [DKSF04]. However, Dohle *et al.* found activations mainly in primary and higher-order visual areas. The discrepancy can be explained by the fact that Dohle's

subjects had to perform a precision task between thumb and the index finger, hence, had to decode the body configuration from a visual image. In the present study, by contrast, subjects were required to perform circular movements that required the decoding of the position of the fingertip and translating it into the appropriate motor command. As indicated by the differences in kinematic metrics, this is performed solely due to the laterality of the visually presented arm.

The activated precuneus region has been associated with the area V6 / V6A in non-human primates. This area was extensively researched by Galletti *et al.* [GFGK99, GFKG99]. Especially the V6A area was linked to visual guidance of arm movements [FGKG01, GFKB97]. Other studies revealed the importance of area V6A for the integration of visual signals, as well as eye and arm movement [BFM$^+$00, BFG$^+$01]. V6A has been discussed as an ideal candidate for the early coding of arm movements within the parieto-premotor network subserving reaching movements [GGK$^+$01]. This has been, furthermore, supported by findings of Battaglini *et al.* who showed that lesions to area V6A lead to misreaching [BMS03]. The correspondence of the precuneus and the V6 / V6A area has been corroborated by recent activation studies in humans that confirmed the role of the precuneus for the adaptation of body configuration for the task of reaching a target in space [dJvdGP01, CGW$^+$02]. This area is also activated during the interpretation of visually presented hand configurations (when compared to body orientation) [VdLV$^+$02]. Lesions of this area are thought to cause the clinical picture of optic ataxia, the deficit in reaching towards visually presented targets in spite of unimpaired motor and sensible functions [PV88, BFM$^+$98].

Our results and these findings allow for the conclusion that the precuneal region in humans provides the anatomical substrate for the integration of visual and kinesthetic information regarding arm movements. As already mentioned in the previous discussion, the virtual arm avatar was capable of inducing strictly contralateral activations similar to those reported in studies using real stimuli, only. This allows for the conclusion that a virtual arm avatar moving synchronously to the real arm is capable of inducing visuo-motor perceptual processes and is, therefore, perceived and accepted as the own body.

4.3.3.3 The virtual avatar is integrated in the body schema

The key question of our experiment clearly was the issue to what extent the computer-mediated virtual arm avatar is integrated into the body schema, i. e., whether it is accepted as the own body.

In the subjective reports given all participants of the study recognized the virtual avatar as a human arm and did not experience any discontinuity in the perception of the kinematic mapping. To which extent the subjects really believed that the virtual arm was their own limb, however, can, of course, not be delineated therefrom.

The use of virtual models of the human body for neurophysiological studies has recently been challenged. In a PET study, Perani and colleagues compared the

observation of real hand movements and those of virtual models in two different levels of abstraction. They found different, however, overlapping cerebral activations [PFB+01]. The virtual models were represented in an immersive way. Perani concluded that computer generated models of the human body should only be used carefully for psychophysical experiments. This interpretation was weakened by a recent study by Farrer *et al.*. When they examined cerebral activation during monitoring of a virtual hand moving synchronously to one's own hand with a variable orientational mismatch to the actual hand, they found a significant variation, i. e., an increase of activation in the inferior parietal lobule and the inverse relation for the right posterior insula in conditions referring to the experience of agency [FFG+03]. Importantly, these areas showed no significant activation in a control condition that was a pure observational task. Thus, Farrer's work can be interpreted such that virtual representations are recognized as parts of the own body as soon as they move synchronously to one's own movements.

Our findings allow an analog interpretation. The parameter estimates in the precuneal region show a variation of activity across the experimental conditions that can not be explained by a direct two-factorial design (*cf.* Fig. 4.5 right column). The largest difference in activation can be found for the two canonical conditions RA– and LA+. The two non-canonical conditions RA+ and LA– clearly cause less change in activation. Therefore, would the virtual arm be simply recognized as an abstract object it would result in activations following a strict two-factorial design. As this is clearly not the case, it can be concluded that the virtual arm is analyzed in a way similar to the own body limbs thereby further justifying the use of virtual body representations for psychophysical experiments given that synchronous motion tracking is used.

Even though the degree of variation between the perception of the own body and a virtual representation, i. e., an avatar, is not known in full detail, and despite the fact that pure observation of real *vs.* virtual objects cause differences in activation, it is apparent, that virtual objects moving synchronously to the own body cause significant cerebral activations, which can only be explained through an integration of the perceived virtual avatar in the mental representation of the own body. In summary, the discussion allows for the conclusion that the virtual avatar is perceived as the own body.

4.3.4 Evaluation of the Hypotheses

The aim of this study was to clarify the integration of a virtual avatar in cerebral processes attributed with perception of the own body and, therefore, the extent of identification of the avatar with the own body. As argued in the last section, the results support the conclusion that, in fact, a virtual body representation is perceived as the own body as long as synchronous kinematics is used. The particular contribution of this study to working hypotheses of this work is the following:

Kinematics. This study clearly indicated synchronous kinematics to be essential for the acceptance of the avatar as the own body. Furthermore, the avatar was capable of inducing cerebral activation in regions that integrate visual and kinesthetic information. Given the fact that an avatar with synchronous kinematics is capable of inducing these activations, and given the definition of presence as suspension of disbelief towards the substitution of reality by media, there is strong evidence for a positive influence of synchronous kinematics on subjective preference.

Presence. The study clearly supports the aforementioned theory of Biocca regarding the interconnection between avatar and body schema in virtual environments (*cf.* section 1.3.3). As this study was performed in an HMD scenario, it does not allow for direct conclusions on presence in room-mounted virtual environments. However, one fact is of importance: even though a first person perspective was used to observe the virtual arm in the HMD the avatar was displayed with a particular offset. This fact allows for the assumption that similar cerebral processes are induced if avatars in room-mounted virtual environments are used.

Usability. In view of the usability hypothesis the study revealed some important facts regarding precision, i. e., the form of motion. It has been shown that the visual percept significantly alters kinematics. The motion of the (real) body was optimized in a way that led to an optimized energetic effort in the kinematics of the observed (virtual) body. This allows for the assumption that avatars mediate and alter kinematics in the reality in that it is optimized and adapted to the virtual world. This means that avatars contribute to the achievement of natural, likely more precise kinematic performance in the mediated environment.

4.4 Kinematics Study

The large contribution of synchronous kinematics on the acceptance of the own body has been concluded from the *virtual arm study*. However, the influence on presence has not been clarified. The *kinematics study* presented in this section extends the research into the influence of kinematics on presence, utilizing the technology proposed in chapter 3, most notably the whole-body avatars and synchronous kinematics. In addition, several presence measurement techniques are used that allow for a discussion and comparison of metrics and their sensitivity to presence.

The second focus of this evaluation is the utilization of avatars in room-mounted virtual environments. The user will perceive his avatar on a spatially shifted position and will also be (visually) aware of his own body. The key point in this study is whether a synchronous avatar in a stressful virtual environment is capable of absorbing the body schema, the mental representation of one's own body. As

described in section 1.3.3, more investigations are needed in this particular room-mounted scenario with a doubled visual information, i. e., the perception of the own natural body and the virtual avatar. A discussion of this problem is crucial in the application of avatars viewed from an exocentric perspective that can be found in collaborative environments, teaching and tutoring scenarios, gaming applications, as well as in complex working environments such as surgical simulators and industrial assembly systems. The results of this study are anticipated to contribute to the discussion on the design and the usability of virtual environments.

In this study, impact of avatars with different kinematic simulations is measured. Although the measurement of presence remains a subject of controversy (*cf.* section 1.3.4), two presence measures are being applied: a subjective questionnaire and an objective physiological measures. As described in previous sections, objective data related to presence are most likely to be measured in stressful virtual environments. Therefore, a sports application in the form of a virtual snowboard ride has been chosen.

4.4.1 Methods

Participants. A total of 27 normal subjects participated in the study (14 males and 13 females, age: 19 to 32 years, mean: 23.6 years). 24 subjects were right handed as assessed with the Edinburgh handedness inventory [Old71]. Subjects were asked about their experience in riding a snowboard to estimate the foot to be placed in front of the board, i. e., the footedness. Those who could not determine their footedness were asked to perform a test: they had to run approximately 2 meters, jump in the air, and land sideways turned by 90°. The same orientation was afterwards used on the board input device. A total of 66.7 % of subjects used the left foot in the front (so-called regular stance, mostly favored by right-handers). In 44.4 % of cases the footedness did not correspond to the handedness, i. e., right-handed subjects used the right foot in front (and *vice versa*). All subjects had normal or corrected-to-normal vision (acuity both eyes far $\geq$ 20/30, acuity both eyes near $\geq$ 14/14 Sneller equivalents), as measured with the Titmus vision tester (A. Zeiss, Petersburg, Virginia). Binocular ability ranged between 30 % and 95 % on the Shepard-Fry scale (mean: 88.7 %). One subject with binocular ability at 70 % reported difficulties with 3D vision. One subject developed strong symptoms of simulator sickness (nausea) after completion of all trials. Four subjects reported vertigo and traces of nausea during the experiment, which, therefore, has been immediately aborted. The subjects were not previously briefed about the purpose of the study.

Design. Subjects were asked to perform a snowboard ride on a virtual slope. The snowboard was steered by shifting the weight to the left or to the right side of the input board. The ride was accelerated by leaning forward and slowed down by leaning backwards. The subjects were required to ride as fast as possible without hitting

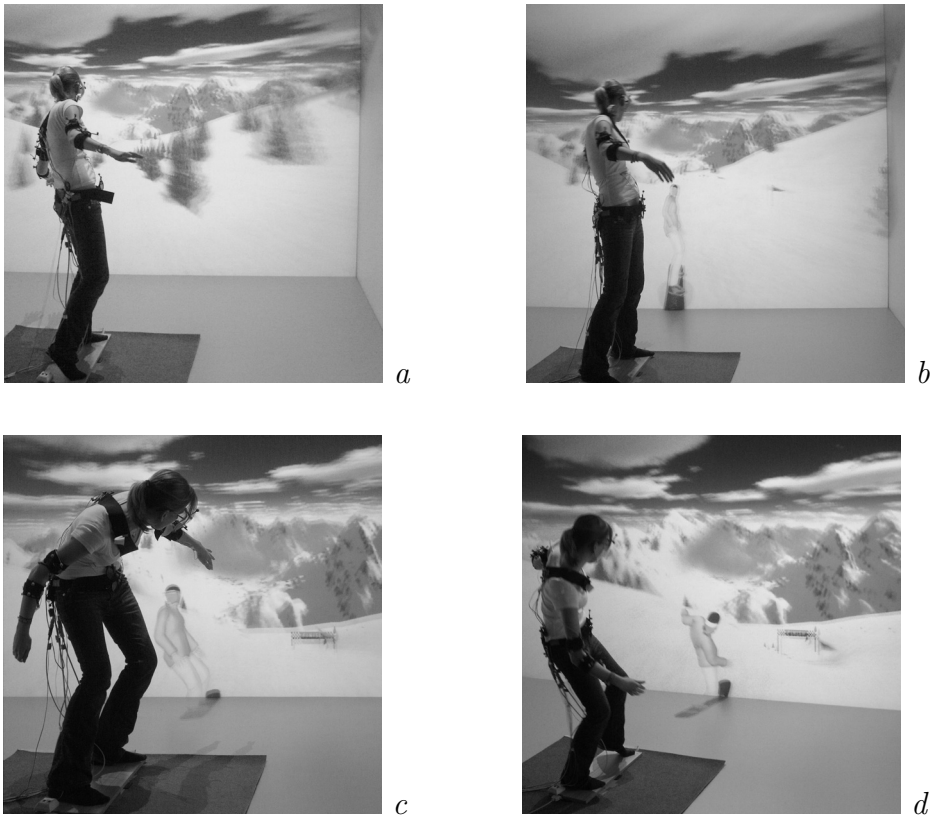


Figure 4.6. Experimental conditions in the *kinematics study*: (*a*) no avatar N–, (*b*) rigid avatar N+, (*c*) an animated avatar A+, and (*d*) a fully tracked avatar T+.

obstacles or leaving the track. Two main factors were manipulated in the study: the avatar viewed from the third person perspective, and the kinematic simulation of the avatar. Either the avatar was displayed (+) or not (–). For the displayed avatar the kinematic simulation was altered from no simulation (N), animation with pre-recorded motion sequences (A), and synchronous full body tracking (T). Therefore, the within-subjects design consisted of four different experimental conditions: N–, N+, A+, and T+ as illustrated in Fig. 4.6.

Procedure. The study consisted of an introductory, training, experimental, and a debriefing session. During the introductory session subjects were presented the hardware and the task of steering an avatar in a sports application. They were informed about simulator sickness and asked to immediately report any occurrence of symptoms. Prior to VR exposition a vision test has been conducted. In order to adapt the stereo rendering individually the eye base has been measured with the PD meter PM-600 (Nidek, Gamagori, Japan). All participants answered a pre-test questionnaire. During the training session subjects have been placed on a virtual snowboard device and provided with optical tracking sensors on their limbs and

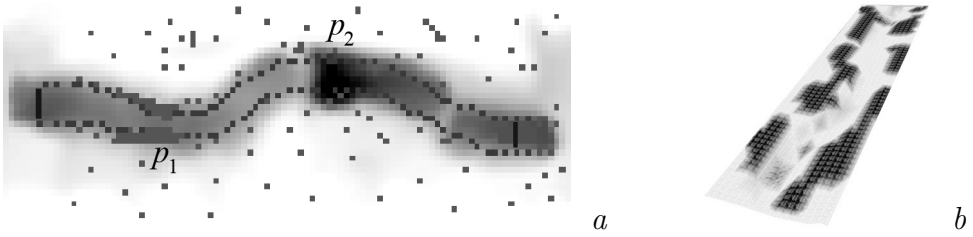


Figure 4.7. Design of the experimental course: (*a*) gray values correspond to the steepness of the slope, red pixels define the position of the slope limits, and green pixels the positions of trees. The start and finish gates are defined as blue lines. p_1 marks the funnel-shaped constriction (FUNNEL) and p_2 the steep part of the slope (JUMP). (*b*) The final rendering of the slope.

electrodes for physiological measures. The subjects were given time to accommodate to the hardware and the application and were asked to perform four test rides under condition T+ followed by four test rides under condition N–. The experimental session included four blocks for the conditions N–, N+, A+, and T+. The blocks have been executed in a pseudo-randomized order and consisted of four trials. Subjects were instructed to ride as fast as possible without leaving the virtual course, or colliding with trees. In addition, they were asked to lift their dominant arm immediately when passing the point of interest JUMP (*cf.* Fig. 4.7). An electrocardiogram, the galvanic skin response, and the expansion of the chest have been measured during the trials. After each block subjects were required to answer an online version of the IPQ presence questionnaire (*cf.* section 1.3.4) on a Samsung Q1-900 tablet PC (Samsung, Seoul, Korea). The debriefing session included a post-test questionnaire. The subjects were asked to remain in the laboratory until they fully re-accommodated to reality.

Apparatus. The experiment has been carried out in a CAVE-like environment with five sides (four walls and the floor) with a resolution of 1600 × 1200 pixels per plane delivered by Sim6 Ultra projectors (Barco, Kortrijk, Belgium). While the back side was left open the stereoscopic images for the remaining four walls were generated by a Linux PC cluster, consisting of eight clients (Pentium 4, GeForce FX6800, 4GB RAM) and a server (Dual Xeon, 4 GB RAM). An optical tracking system consisting of four ARTtrack1 cameras has been applied (Advanced Realtime Tracking, Weilheim, Germany). Clusters of markers have been used to drive the tracking algorithm as described in section 3.2.2.2 with markers on each forearm, each upper arm, the head, the neck, and two on the pelvis. A board input device configured with four weight sensors has been used to steer the virtual snowboard and the avatar. The physiological measures have been retrieved with the BioSemi biofeedback device (BioSemi, Amsterdam, Netherlands). Two electrodes have been fixed on the thenars of the non-dominant hand with a distance of ≈4–5 cm from each other to measure the galvanic skin response; the electrocardiogram was retrieved

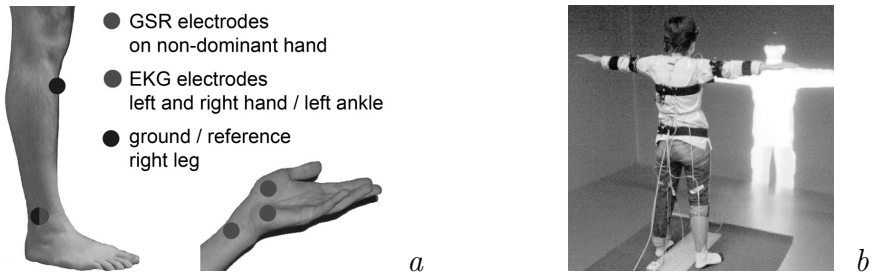


Figure 4.8. (*a*) Electrode set-up for measurement of physiological reactions. (*b*) A participant configured with electrodes and optical markers during calibration of the tracking system.

with three electrodes: one on the left, one on the right lower forearm, as well as an electrode on the left ankle; the expansion of the chest during respiration has been retrieved with a flexible belt around the thorax. The device uses two electrodes for normalization / grounding of the measured data: an active electrode is placed on the right ankle, and a passive electrode is fixed atop of the right upper tibia (*cf.* Fig. 4.8). All physiological measures have been recorded for offline analysis. In addition, the BioSemi system recorded a set of triggers that have been manipulated by the application.

The environment was run by a NeuroMan application (*cf.* section 4.2) based on the ViSTA software utilizing the OpenSG library. In addition, the software used code from the TuxRacer project[2] for physics simulation and the game engine. The course was designed as a slope of 400 m length, started and ended with gates, and was marked with flagged poles on each side. It, furthermore, contained two points of interest: a funnel-shaped constriction made of tree obstacles (FUNNEL), and a surprisingly steep part of the slope that mostly required a jump (JUMP) as depicted in Fig. 4.7. The application generated triggers for the BioSemi device for all events.

The avatar was displayed and simulated with the virtual humanoid toolkit VRZula. The virtual humanoid was defined as an H-Anim model with joints according to the level of articulation one. The geometry consisted of 8,884 vertices and was displayed with VRZula using GPU-based vertex blending.

In all conditions the application was steered with the board input device. In condition N+ the orientation of the rigid avatar was adapted to the current curvature of the slope. To realize the animated condition A+, five motion sequences have been recorded: neutral standing, acceleration, slowing down, turning left and right. These sequences have been connected with motion graph techniques described in section 3.2.2.3. Data from the board input device were used to synthesize the motion of the avatar from the motion graph. Finally, the avatar was tracked by the motion of the body of the user in condition T+.

[2]http://tuxracer.sourceforge.net, last visit May, 2007.

Data Processing and Analysis. For each subject continuous physiological data have been measured in 24 trials (two blocks in the training session, and four blocks in the experimental session, each containing four trials), altogether 5 channels and 16 triggers at a sampling rate of 2,048 Hz. The average length of a trial was 30.6 s ($\pm$4.5). The following data channels have been measured:

- three channels for the electrocardiogram,
- the galvanic skin response,
- the expansion of the chest for measurement of respiration, and
- triggers indicating the start and the end of the course, reaching the points of interest, leaving the course, colliding with tree obstacles, and jumping.

The electrocardiogram channels have been used to compute the heart rate. An approach for detection of R waves similar to [PT85] has been applied. This method is intended for real-time detection and, hence, needs a final integration step. The integration has been replaced by a search for local maxima yielding the peaks of the R waves, resulting in the intervals between successive heart beats, the so-called inter-beat intervals. The galvanic skin response channel delivered the current resistance of the skin surface ρ and has been used to compute the skin conductivity; however, most of the data have been biased due to the motion of the subjects and was, therefore, not submitted for statistical analysis. The continuous data obtained from the expansion of the belt around the chest have been used to compute the breath rate through application of a local maximum search algorithm on the bandpass filtered data. The continuous data have been analyzed in two ways, as a function of the time, and as a function of the position on the slope. The latter approach was used when comparing the data for the particular points of interest.

In addition to physiological values the following performance data have been recorded:

- the time needed to complete the rides,
- collisions with tree obstacles,
- position and orientation for all optical tracking markers, and
- data from the board input device.

Verbal remarks expressed by the subjects have been recorded, as well.

Multiple questionnaires have been used to gather further information, like demographic data and experience with VR and sports (pre-test questionnaire), and personal preferences concerning the conditions, as well as feedback (post-test questionnaire). The subjects completed the IPQ presence questionnaire after each one of the four blocks in the experimental session. The presence score has been computed according to [SFR01].

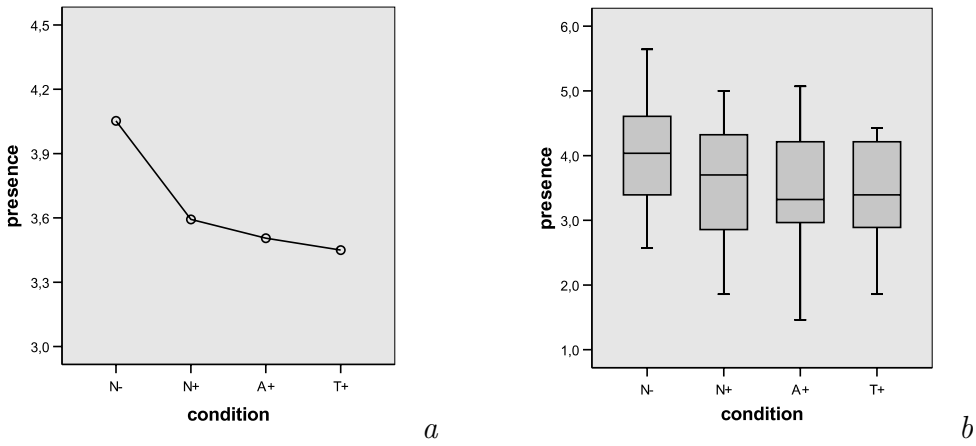


Figure 4.9. (*a*) Estimated marginal means, and (*b*) box plots of presence scores.

Continuous data have been analyzed with Matlab (The MathWorks, Natick, MA) and customized modules. Statistical analysis has been performed with SPSS (SPSS Inc., Chicago, IL).

Data obtained from the five subjects that reported symptoms of simulator sickness and data from one subject that reported serious difficulties with 3D vision have been excluded from the statistical analysis. Due to technical problems the physiological data for one subject have been biased and, therefore, falsified. The presence questionnaire data for one subject have not been saved in the data base. These data have not been used in the corresponding analyses, as well. Altogether, data from 20 subjects have been used for further analysis.

4.4.2 Results

Presence Scores. The presence scores have been analyzed with a one-way ANOVA with repeated measures. A significant main effect due to the subjective estimates of presence was observed. The reported mean presence score was higher in the N– condition ($M = 4.052 \pm 0.186$) than in the conditions using an avatar: N+ ($M = 3.593 \pm 0.205$), A+ ($M = 3.505 \pm 0.186$), and T+ ($M = 3.450 \pm 0.167$). This difference was significant: $F(3, 57) = 5, 36$, $p = 0.003$. A *post hoc* pairwise comparison of all conditions revealed that condition N– significantly differs from all other conditions at the $p < 0.05$ level, and that presence estimates for conditions N+, A+, and T+ can not be significantly separated from each other (*cf.* Fig. 4.9).

Physiological Measures and Performance. The repeated measures one-way ANOVA showed no significant effect between the conditions neither for the heart rate ω^{HR} ($F(3, 57) = 0.423$, n. s.), nor for the mean breath rate ω^{BR} ($F(3, 54) = 0.281$, n. s.).

Table 4.4. Mean values and standard deviations for physiological and performance measures: mean breath rate ω^{BR}, mean heart rate ω^{HR}, change in the heart rate $\Delta\omega^{\mathrm{HR}}$ at points of interest p_1 (FUNNEL) and p_2 (JUMP), the average speed v and the achieved maximal speed $v_{\max}$, time for completion of a ride t, as well as the mean number of collisions n_{col} and jumps n_{jmp} per trial given for each experimental condition.

Measure	Condition			
	N–	N+	A+	T+
presence	4.05 (± 0.833)	3.59 (± 0.917)	3.51 (± 0.830)	3.45 (± 0.748)
ω^{BR} / min^{-1}	21.95 (± 3.118)	22.01 (± 3.575)	22.00 (± 3.189)	22.35 (± 3.346)
ω^{HR} / min^{-1}	104.3 (± 11.17)	104.6 (± 15.11)	103.2 (± 11.58)	104.6 (± 11.93)
$\omega^{\mathrm{HR}}_{p_1}$ / min^{-1}	103.6 (± 11.17)	103.4 (± 14.63)	103.0 (± 11.60)	104.1 (± 12.08)
$\Delta\omega^{\mathrm{HR}}_{p_1}$ / min^{-1}	0.641	0.569	0.931	0.756
$\omega^{\mathrm{HR}}_{p_2}$ / min^{-1}	104.6 (± 11.70)	103.8 (± 14.61)	103.8 (± 11.68)	104.7 (± 12.41)
$\Delta\omega^{\mathrm{HR}}_{p_2}$ / min^{-1}	1.056	0.486	0.822	0.803
v / m/s	15.04 (± 1.229)	14.99 (± 1.022)	14.86 (± 0.895)	14.68 (± 1.177)
$v_{\max}$ / m/s	21.85	22.18	22.01	21.78
t / s	30.49 (± 4.213)	30.09 (± 2.773)	30.86 (± 2.556)	31.44 (± 3.619)
n_{col}	0.97 (± 0.776)	0.87 (± 0.800)	1.05 (± 0.740)	1.06 (± 1.208)
n_{jmp}	5.24 (± 0.712)	4.79 (± 0.658)	4.96 (± 0.943)	4.71 (± 0.766)

The use of an avatar had no significant influence on the performance metrics. While the amount of collisions with trees n_{col} did not change over conditions ($F(3, 60) = 0.209$, n. s.) there is a (non-significant) tendency in the amount of jumps n_{jmp} during a ride ($F(3, 60) = 2.584$, $p = 0.062$). Subjects jumped 5.24 times per trial on average without an avatar compared to 4.82 jumps in conditions presenting an avatar. Yet another tendency was observed in the mean speed v ($F(3, 60) = 1.728$, $p = 0.160$) and the time t ($F(3, 60) = 1.757$, $p = 0.165$). Subjects were slightly faster with an avatar (30.49 s at an average speed of 15.04 m/s) than without (30.80 s at 14.84 m/s).

All values for the measured physiological and performance metrics are given in Table 4.4.

Subjective Reports and Preferences. Mainly, the subjects did not comment their performance during the trials, however, eleven have expressed surprise by verbal remarks, mostly during the training session, after the switch from training condition T+ to N–. The remarks—originally in German—have been as followed:

- "Where am I?" (expressed independently by three subjects)
- "Where do I see myself?"
- "I don't see myself. Is that right?"
- "I need the guy in order to steer!"
- "I think I'll perform worse without [*the avatar*]."

In contrast, a subject clearly asked if the avatar could be switched off again after completing the training with condition N– and starting the experiment with block A+. Another subject was suddenly scared during the experimental session while switching from block T+ to condition N– and asked to be warned in advance if the avatar will be switched off again. Interesting behavior has been observed in two subjects: the first apparently enjoyed interacting and playing with the avatar by moving his limbs in condition T+. The second observed his own arm doubtfully during the kinesthetic–optical mismatch that occurred when the subject lifted his arm without optical feedback at the JUMP point of interest, after completing condition T+ and switching to A+.

After completion of the experimental session subjects were asked to estimate the condition they preferred most. On average, 66.7 % of subjects preferred the N– condition over the conditions with an avatar N+, A+, and T+. In conditions presenting an avatar the choice was 29.6 % on average for the tracked condition T+, 2.2 % for A+, and 1.5 % for N+, only. The results are presented in Fig. 4.10.

Correlation between Objective and Subjective Measures. A significant negative correlation was found between the reported presence scores obtained by questionnaires and all measured heart rates, i. e., for the mean heart rate ω^{HR} ($r(79) = -0.332$, $p = 0.003$), the value at the FUNNEL point of interest $\omega^{\mathrm{HR}}_{p_1}$, ($r(78) = -0.349$, $p = 0.002$), and at the JUMP point of interest $\omega^{\mathrm{HR}}_{p_2}$, ($r(78) = -0.334$, $p = 0.003$). A gender-specific analysis of the mean heart rates revealed even a stronger correlation with presence for males ($r(39) = -0.484$, $p = 0.003$) than for females ($r(43) = -0.313$, $p = 0.039$). By contrast, no correlation was found between the measured breath rate and the presence scores ($r(79) = 0.081$, n. s.). The data are presented in Fig. 4.11.

The analysis of the performance metrics yielded two significant correlations with presence scores: a positive correlation between the mean speed v and the reported subjective presence score, $r(83) = 0.250$, $p = 0.023$ and a negative correlation between the mean time t and presence, $r(83) = -0.217$, $p = 0.048$, It has to be

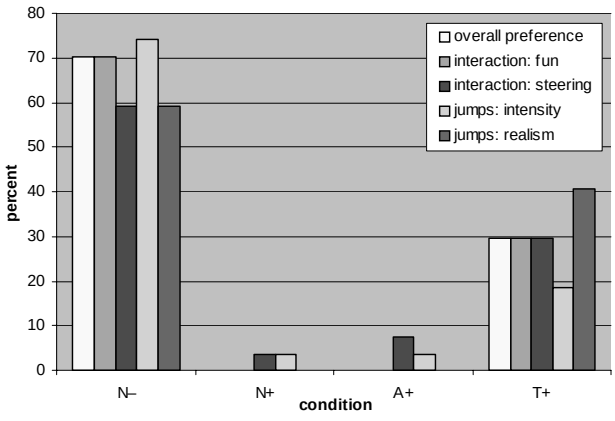


Figure 4.10. Personal preferences as estimated by questionnaires after completion of all experimental blocks. Subjects were asked to choose the most preferred condition with regard to personal preference, interaction and the experience of jumps on the virtual track.

noted that even though both values t and v are strongly correlated, $r(84) = -0.962$, $p < 0.001$, they are independent of each other, because subjects traveled different distances during trials. While several subjects preferred to ride straight ahead, others made extensive use of the steering possibilities and produced tracks with a higher curvature and, therefore, longer distance.

4.4.3 Discussion

The results of the *kinematics study* clearly show that while the representation of an exocentric avatar decreases presence in room-mounted virtual environments, body tracking is subjectively preferred over animated or rigid avatars. The study also revealed a correlation between presence scores and physiological measures, as well as between presence scores and performance metrics.

4.4.3.1 Identification with exocentric avatars is not persistent and, therefore, decreases presence

The reported presence is strongly dependent on the use of an exocentric avatar. The score is significantly higher in the condition without an avatar N– than in all conditions using variations of the exocentric avatar. Room-mounted virtual environments—as used in the study—strongly immerse the user within a virtual world without restricting the view on his own body. The application of an avatar in such a configuration is, therefore, bound to mental processes that resolve the identification of the body schema with the two competing body representations. Evidence for this mental process has been found in the comments of subjects on the avatar principally referred to as oneself ("Where am I?", "Where do I see myself?", *etc.*) that were mostly expressed during the training session. However, the decrease in the

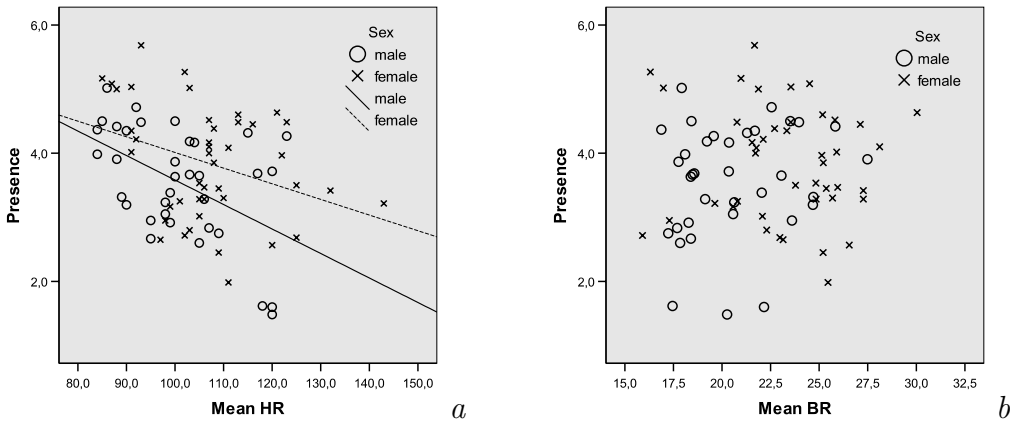


Figure 4.11. Correlation between (a) the mean heart rate, and (b) the breath rate with the reported presence scores. The graphs show gender-specific subgroups. The best fitting lines are displayed for the heart rate, which is significantly correlated, $r(79) = -0.332$, $p = 0.003$. The breath rate and the presence scores were not correlated, $r(79) = 0.081$, n. s.

subjective presence for conditions presenting an exocentric avatar indicates that the avatar did not attract the user's body schema permanently and was, therefore, perceived as disturbing in the experience of the virtual world. Apparently, the avatar was not permanently preferred as a mean for identification within the virtual world, contrary to the conclusions in [Tay02]. Our findings allow an interpretation in terms of the results by David *et al.* who investigated the dependence of perspective on the quality of agency: avatars viewed from a more exocentric perspective induced a lower quality of agency than egocentric avatars [DBC+06]: agency can be seen as a consequence of the identification of the body schema with the avatar. In contrast to the presence scores obtained, no significant differences have been found for the physiological and the performance measures, implying that subjects performed equally well in all conditions. However, the tendencies observed for changes in the heart rate at the FUNNEL point of interest, as well as the amount of jumps, and the mean speed allow the conclusion, that subjects tend to perform better in condition N–. This implies, that more attention has been paid to the completion of the task in condition N– and that, presumably, attentional resources have been allocated in the remaining conditions for the mental processes involved in identification with the avatar.

By way of an overall conclusion, the effect observed proved to be not strong and persistent enough for creating a significant identification with the avatar albeit a mental binding of the body schema to the avatar took place. Therefore, the reported presence was lower in conditions involving an avatar compared to the condition without one.

4.4.3.2 Synchronous kinematics is the preferred choice for avatars

While the analysis of the subjective presence scores has not revealed statistical differences among the conditions using an avatar: N+, A+, and T+, subjective preference estimates clearly indicate T+ to be the favorite condition. On average $1/3$ of subjects have identified with the exocentric avatar and have, therefore, deprecated the condition N–. Among those subjects 88.9 % preferred the tracked exocentric avatar. Going more into detail, only 4.4 % and 6.7 % of these subjects preferred the condition N+ an A+, respectively when asked about the interaction and steering, as well as the intensity of jumps. It is most likely that these subjects have identified with the avatar on the one hand, yet have chosen among the conditions that do not attract the proprioceptive attention on the other hand. Surprisingly, the clear preference for T+ was not reflected in the presence scores obtained.

It is safe to conclude that an important number of users prefer to identify with an exocentric avatar whose motion is directly bound to the user's body. Furthermore, it is noteworthy that this identification does not lead to an important negative influence on user performance. Therefore, synchronous kinematics is capable of ameliorating the overall usability of a system that uses avatars.

4.4.3.3 Increased performance in higher presence scenarios

As mentioned in the last section, presence is regarded as a factor of usability. To date, task and system performance are used to assess the usability of a system. The results revealed a correlation between presence scores and task performance. In particular, a correlation has been found for the mean speed v and mean time t. Subjects performed better with increasing presence, i. e., they have been faster and they needed less time to finish the task. This correlation has been further substantiated by the intra-individual analysis of the four conditions where (non-significant) tendencies have been found for v, t, and, in addition, the number of jumps n_{jmp}. The interrelationship between presence and performance leads to the conclusion that higher presence is perceived in systems with higher usability.

4.4.3.4 The ambivalent role of the heart rate in estimating presence

In view of the estimation of presence by objective measures, such as physiological metrics, the study revealed several significant results. Most importantly, it has been found that the subjective presence score correlates with the heart rate in an inter-individual comparison. This correlation was continuously negative for the heart rate measures ω^{HR}, $\omega_{p_1}^{\text{HR}}$, and $\omega_{p_2}^{\text{HR}}$. It has been further substantiated by a gender-specific analysis. The correlation implies that with an increased presence score the heart rate slowed down. Heart rate has been shown to be an objective measure for presence, e. g., in [MIWB02]. Meehan *et al.* conclude that higher presence is bound to a higher

heart rate. They argue that "higher presence" virtual environments should evoke stronger physiological reactions because they are closer to reality. The difference in Meehan's studies compared to our work is the comparison of virtual environment set-ups that induced stress in the users dependent on their degree of realism. This is achieved by stimuli for additional modalities (like haptics) and by variation of the content of the virtual world. By contrast, our study used one task (snowboard ride) in one specific virtual environment that induced a certain amount of stress and physiological reaction. The measured heart rate was, therefore, dependent on the variations in the application and kinematics of the exocentric avatar. As concluded previously, the decrease in the reported subjective presence scores was due to a non-persistent identification with the avatar, which was perceived as perturbing. The additional stress induced by the concurring representation of the exocentric avatar explains the increase in the measured heart rate combined with a decrease in the presence scores.

In summary, stressful situations or environments are capable of inducing different sensations of presence, as well as different pace in the heart beat. However, while comparing the presence scores and physiology measures, the cause for variation in both has to be analyzed closely. Virtual environments whose degree of realism is directly related to the degree of stress will lead to an increase in both measures: a higher presence and a higher heart rate in a more realistic scenario. In our case, a conflict in identification with an exocentric avatar led to increased stress and, therefore, an increased heart rate in spite of reduced presence. To conclude, if the relation between realism and the induced stress is known for a certain virtual environment scenario, then the heart rate can be used as an objective measure for presence.

4.4.4 Evaluation of the Hypotheses

The *kinematics study* aimed at evaluating the influence of exocentric avatars on the sensation of presence, and at revealing interrelations with usability. It has shown that presence is highest if the center of action is the user himself, i. e., if no avatar is used. Even though users did identify with an avatar, this effect was not persistent. Avatars with synchronous kinematics were the preferred choice in conditions using an avatar. Furthermore, the study revealed an interrelation between physiological and performance data and subjective presence estimates, suggesting that the heart rate can be used as a measure for presence, and that presence affects usability to a certain extent. The contributions to the working hypotheses are as follows:

Kinematics. The kinematics hypothesis is supported by the subjective preference scores obtained in this study where nearly all participants have chosen an avatar with synchronous kinematics.

Presence. The main finding of this study partially supports the presence hypothesis. It has been shown that presence is highest in room-mounted virtual en-

vironments when the user directly interacts with the virtual world, i. e., when no avatar is displayed. The usage of an exocentric avatar led to a decrease in the presence scores. However, subjective reports showed that users strongly identify with the avatar. As discussed, this effect has not been transient. The processes involved in the resolving of the identification dilemma of the body schema with the two concurring body representations created a conflicting situation that caused a decrease in presence.

Usability. The influence of presence on usability is clearly shown by the obtained correlation between presence scores and performance metrics. This study indicates that higher presence goes hand in hand with better performance that is usually achieved in systems with higher usability.

4.5 Did the Working Hypotheses Hold their Promise?

This chapter has presented hypotheses and studies exploring the influence of avatars on presence in virtual environments. The hypotheses address topics of synchronous kinematics, usability, presence, and avatars in room-mounted virtual environments.

The *virtual arm study* has revealed the influence of synchronous kinematics on the acceptance of avatars as body representations in virtual environments. The *kinematics study* has clarified the influence of exocentric avatars used in room-mounted virtual environments on the sensation of presence. Both studies deliver a broad foundation for conclusions and suggestions on the application of avatars in virtual environments.

It has been shown that avatars with synchronous kinematics attract the body schema and lead—at least temporarily—to an identification with the virtual body representation. This identification is strong enough to influence the kinematics of the user as shown in the *virtual arm study*. Yet it has also been shown that exocentric avatars in virtual environments, in general, induce mental processes that allocate attentional resources needed to resolve the identification with the two concurring body representations, leading to a decrease in presence and performance. It is this interrelationship between the identification with the virtual avatar and the additional mental processes that bears the potential in application of virtual humanoids as avatars: on the one hand, the identification may strongly be used to positively influence the user; on the other hand, it can slow down the working process if used inappropriately.

4.5.1 Evaluation of the Hypotheses

The studies presented in this work allow for the following evaluation of the initial working hypotheses.

Both studies support the kinematics hypothesis, that claimed synchronous kinematics to be capable of increasing subjective preference for an avatar. While the *virtual arm study* showed that synchronous kinematics is essential for the identification with an avatar, the second study substantiated that users do prefer fully tracked avatars over animated or rigid virtual humanoids.

The presence hypothesis was partially approved. Both studies clearly showed the occurrence of mental processes involved in the identification of oneself—i. e., the body schema—with the avatar. Yet the studies have not confirmed the postulated increase in presence. As discussed, the identification with an exocentric avatar in room-mounted virtual environments is not persistent and stable. Therefore, the co-presence of the avatar creates an unfamiliar situation that, finally, leads to a decrease in the subjective presence scores.

Given the characterization of usability by its factor task performance, then the usability hypothesis is supported by the two studies. The first study showed the capability of avatars to modify kinematics of the user in that the final motion in the virtual environment was optimized. This also implies that the virtual task is achieved with a higher performance if mediated by a virtual avatar. The discussed connection of avatars and presence leads to the conclusion that presence affects performance and, therefore, usability. In addition, the *kinematics study* revealed a direct correlation between performance and presence metrics. It can, therefore, be concluded, that presence influences usability of virtual environments.

In addition to the evaluation of the hypotheses, the results contribute to the understanding of capabilities of avatars in room-mounted virtual environments; an open issue imposed by current technology itself. The main feature of room-mounted virtual environments is the fact that the user is directly immersed in the virtual environment and constitutes the center of action similar to real world interaction. Obviously, egocentric avatars can not be realized in room-mounted virtual environments, because of the perturbing interference between the real body and the virtual avatar representation. Even though the *kinematics study* found a decrease in presence during display of avatars it also confirmed the strong identification with the exocentric avatar. Therefore, the choice for an exocentric avatar in room-mounted virtual environments is a classical trade-off decision: on the one hand, the avatar representation competes with the real body in the process of identification with the body schema. The unresolved and altering situation causes a decrease in presence. On the other hand, the exocentric avatar is likely to be capable of, at least temporarily, attracting the body schema and, therefore, influencing particular performance, like kinematics as shown in the *virtual arm study*. For this reason, application scenarios and tasks in room-mounted virtual environments have to be further analyzed to determine the benefits of an application of avatars.

Table 4.5. Application scenarios for avatars in virtual environments and their effects on users. Scenarios are grouped according to the main interaction task of manipulation and navigation. Tasks marked with # are direct results of the studies presented.

	Influence		Task	
Configuration	Identification	Presence and Usability	Manip.	Navig.
head-mounted display				
no avatar	–	missing frame of reference	−	∘
egocentric	persistent	alters kinematics	++ #	+
exocentric			+ #	+
room-mounted virtual environment				
no avatar	–	user as center of action	∘	+ #
exocentric	temporal	concurring body representations	+	∘ #

4.5.2 Analysis of Application Scenarios for Avatars

As mentioned previously, most research on avatars and presence has been carried out with HMDs. This work has contributed a study on the influence of avatars in room-mounted virtual environments, which require an exocentric representation of the virtual humanoid. As exocentric avatars interfere with the real body in room-mounted set-ups, their application has to be considered in view of a possible allocation of attentional resources. The subsequent discussion of application scenarios is structured in the two main tasks of manipulation and navigation (*cf.* [BKLP05]). In this work selection and orientation are both discussed as parts of manipulation and navigation, respectively.

HMD set-ups are the most beneficial application scenario for egocentric avatars. The avatar creates a frame of reference and identification for the user within the virtual world. Manipulation and navigation are, therefore, well supported by an egocentric avatar. Manipulation tasks performed in HMDs without an avatar are more difficult because the visual stimuli matching the direct proprioceptive feedback are missing. Navigation without an avatar can be done relatively simply mainly dependent on the chosen metaphor and behavior for the navigation. However, the lack of a frame of reference can predispose symptoms of simulator sickness. Although David and colleagues found a link between agency and the chosen perspective, i. e., exocentric *vs.* egocentric avatar [DBC+06], the *virtual arm study* indicates that identification of the body schema occurs even in exocentric avatars, which are in a certain proximity to the user's body. In addition, it has been shown that synchronous kinematics in the avatar influences the user's kinematics resulting in optimized performance in the virtual world. The characteristics for manipulation and navigation are, hence, comparable to egocentric avatars.

Room-mounted virtual environments are broadly used and are the current state of technology in VR. The user represents the center of the virtual world and is surrounded by computer-generated stimuli. Therefore, interaction in such environments is close to interacting in reality. This is, in particular, the case for navigation. However, manipulation entails several difficulties, the most prominent being the occlusion problem in direct manipulation of items. As soon as the user touches or reaches for a virtual object perceptive problems occur because of the mismatched accommodation of the eyes. This drawback can be resolved by interaction metaphors or indirect interaction techniques. As mentioned before, only exocentric avatars can be used in room-mounted set-ups appropriately. Exocentric avatars have been shown to create a temporal identification with the user's body schema, and could, therefore, be used to realize optimized occlusion-free manipulation. However, as shown in the *kinematics study* avatars in room-mounted virtual environments attract attention and may decrease presence and performance. Therefore, the manipulation task has to have an important amount of kinematic complexity that is beneficially influenced by the avatar. As manipulation in room-mounted virtual environments has always to be done from a certain distance in order to avoid the occlusion problem, an avatar can, in addition, serve as a frame of reference for the size of virtual objects during manipulation. The allocation of attention and other mental processes makes navigation with an avatar less efficient, i. e., the attention is turned towards the avatar instead of the navigation task itself. For this reason, avatars should be applied in navigation tasks that require the attention to be turned towards the avatar. This is for example the case when avatars create a link to the virtual world, or when they are used to represent additional information, like maps, charts, or hints for further interaction possibilities in the virtual world. Similar to the occlusion problem in manipulation, a further promising scenario is the avoidance of collisions with virtual objects during navigation. The avatar is used as a frame of reference for the user, navigates according to the user, however, avoids unnatural collisions with virtual objects. This navigation metaphor is likely to reduce collisions within the virtual world.

In summary, the results of this thesis suggest the use of virtual humanoids as avatars in several beneficial scenarios, such as most manipulation and certain navigation tasks (*cf.* Table 4.5). If used, avatars should clearly employ techniques for synchronous kinematics as they are the preferred choice in most users and the reason for identification with the avatar.

This chapter has presented studies aimed at clarifying the effects of the application of avatars on the mental processes of a user during perception of and interaction with the virtual environment. The initial working hypotheses regarding preference for synchronous kinematics in avatars and interconnection between presence and usability have been confirmed, yet a presumed interrelation between avatars and the increase in sensation of presence has not been observed. Nevertheless, the results lead to important recommendations towards the improved application of virtual humanoids as avatars in virtual environments.

5
Summary and Conclusions

Virtual humanoids and their application as avatars have been the focus of this thesis. State-of-the-art methods, a comprehensive software system for virtual humanoids, and studies focusing on avatars have been presented and discussed in great detail. As outlined in the Introduction and, in detailed form, in the Motivation, this thesis has specifically addressed two issues, namely:

- Development of a virtual humanoid toolkit for a comprehensive and interactive simulation of virtual humanoids capable of integrating current simulation methods in an open, adaptable framework, enabling further research on interrelations of users and virtual humanoids in virtual environments.

- Experimental evaluation of the influence of avatars on users in room-mounted virtual environments towards further understanding of the influence of synchronous kinematics on presence, and towards clarification of its effects in room-mounted virtual environments.

The goals have been achieved by a number of technical contributions and experimental evaluations.

The thesis contributes the extendible virtual humanoid toolkit VRZula intended for the use in virtual environments. The elementary data structure is based on systems defined in functional anatomy and, therefore, allows for an integral definition and representation of various human body structures. The separation of data and algorithms led to a general framework enabling flexible modifications to simulation approaches, as well as a clear control and propagation of changes even in complex models and distributed virtual environments. Beside the integration of recent simulation methods for visual representation and simulation of kinematics the toolkit contains the following novel approaches:

- The seamless integration of simulation of physiology has been demonstrated on a novel method for interactive, volume-preserving simulation of musculature. This method uses bounding-boxes to compute deformations of muscles in real-time on arbitrary geometries. The representation leads to an improved visual realism compared to established interactive methods.

- Issues of multimodal representation have been addressed by the speech processing approach integrated in the VRZula framework. The system incorporates text-to-speech software and features the synchronized representation of stereoscopic facial visualization and binaural speech rendering.

- Synthesis of motion has been realized with motion graphs. A combination of strategies for the construction of motion graphs has been proposed that yields optimal synthetic motion data with regard to navigation capabilities and degree of realism of animation.

- Body tracking has been solved with an orientation-based approach. This technique has been adapted to the use with optical tracking by adding methods that analyze and evaluate the validity and continuity of data. The overall approach represents a robust method for the tracking of large articulated structures and yields continuous data for the kinematics of the virtual humanoid.

The second major contribution is the experimental evaluation of the usage of avatars in virtual environments. HMDs, as well as room-mounted virtual environments have been researched with studies focused on synchronous kinematics, precision of motion, performance in navigation, and the final impact on presence estimated with subjective, as well as objective measures. The evaluation of hypotheses led to two main conclusions:

- Avatars with synchronous kinematics affect the real motion of the user insofar as the resulting motion of the avatar in the virtual world is optimized.

- The user strongly identifies with an avatar. Mental and attentional resources are bound to the virtual body representation.

These findings have been used in the evaluation of typical interaction scenarios in virtual environments. It has been confirmed, that exocentric avatars in room-mounted virtual environments allocate attentional resources that lead to a decrease in presence. However, the capability of identification and the positive influence on kinematics bear the potential for several gainful application scenarios. As a consequence, the thesis suggests the use of exocentric avatars in room-mounted virtual environments for manipulation in order to avoid the occlusion problem and in navigation if an additional attention and focus on the avatar is desired. This is, for example, the case for avoidance of collisions or display of auxiliary visual information. An additional advantage of such an application of avatars is the frame of

reference established between the user and the virtual world. The usage of egocentric avatars in HMD set-ups has been shown to be advantageous for the majority of interaction tasks. However, these results have been obtained with avatars using synchronous kinematics and the conclusions are, therefore, only valid for this technology, as synchronous kinematics has been shown to be the main reason for the identification and invocation of strong cerebral processes. The thesis, therefore, suggests the use of avatars with kinematics directly derived from the motion of the user.

In summary, the thesis has provided a significant contribution to the simulation of virtual humanoids, and to the understanding of effects of avatars on users, which is of major importance for interaction in virtual environments. The recommendations for application of virtual humanoids in several interaction scenarios ameliorate the working experience and create a frame of identification for the user.

The results lead to some questions that have to be further analyzed in future research, for example:

- The decrease in presence observed in the *kinematics study* is dependent on the immersive capabilities of the room-mounted virtual environment. As this environment already creates a high sensation of presence, it is of interest to compare the change in presence induced by avatars in less immersive systems. Further evaluations are necessary to fully understand the effects of exocentric avatars.

- This thesis researched the factor of kinematics in avatars. As outlined before, the visual degree of realism has interesting effects on the acceptance of the virtual humanoid as an avatar. This effect has been discussed as the "uncanny valley" phenomenon. The studies presented have used a coarse avatar, as well as an avatar with a high degree of realism in appearance. Both avatars have shown their impact on the user. However, further studies that alter the degree of visual realism will have to determine the effects of the "uncanny valley".

The methods for virtual humanoids and the evaluation of their usage as avatars have demonstrated the broad range of application, as well as the benefits brought by this technology. The work presented in this thesis clearly showed that the use of virtual humanoids will significantly improve current and future applications in virtual environments.

Bibliography

[ACP02] Brett Allen, Brian Curless, and Zoran Popović, *Articulated body deformation from range scan data*, Proceedings of SIGGRAPH '02, 2002, pp. 612–619.

[AF02] Okan Arikan and David A. Forsyth, *Interactive motion generation from examples*, Proceedings of SIGGRAPH '02, 2002, pp. 483–490.

[AFO03] Okan Arikan, David A. Forsyth, and James F. O'Brien, *Motion synthesis from annotations*, Proceedings of SIGGRAPH '03, 2003, pp. 402–408.

[AH02] Tomas Akenine-Möller and Eric Haines, *Real-time rendering*, second ed., A K Peters, 2002.

[AHV+05] Claudia Armbrüster, Ines A. Heber, Jakob T. Valvoda, Torsten Kuhlen, Bruno Fimm, and Will Spijkers, *Distance estimation in a VR application: Inter-individual differences and intra-individual stabilities from a psychological point of view*, HCI '05: Proccedings of the 11th International Conference on Human-Computer Interaction (Las Vegas, NV), 2005.

[AKL05] Ingo Assenmacher, Thorsten Kuhlen, and Tobias Lentz, *Binaural acoustics for CAVE-like environments without headphones*, IPT & EGVE '05: Proceedings of the IPT & EGVE Workshop 2005 (Aalborg, Denmark), 2005, pp. 31–40.

[AKLV04] Ingo Assenmacher, Torsten Kuhlen, Tobias Lentz, and Michael Vorländer, *Integrating real-time binaural acoustics into VR applications*, EGVE '04: Proceedings of the Eurographics Symposium on Virtual Environments (Grenoble, France), June 8-9, 2004, pp. 129–136.

[ALP04] Yeuhi Abe, C. Karen Liu, and Zoran Popović, *Momentum-based parameterization of dynamic character motion*, SCA '04: Proceedings of the 2004 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (Grenoble, France), August 27 - 29, 2004, pp. 173–182.

[AN99] Yahya Aydin and Masayuki Nakajima, *Realistic articulated character positioning and balance control in interactive environments*, CA '99: Proceedings of Computer Animation (Geneva, Switzerland), 1999, pp. 160–168.

[And06] Solveyg Anders, *Eurade - Normierung und Evaluation eines Therapiesystems*, Jahresbericht [annual report] 2005/2006, Virtual Reality Center Aachen, RWTH Aachen University, October 2006.

[Ass06] Ingo Assenmacher, *ViSTA - Aktuelle Entwicklungen*, VRCA Workshop 2006 (Aachen, Germany), November 21, 2006.

[AWK+07a] Claudia Armbrüster, Marc Wolter, Torsten Kuhlen, Bruno Fimm, and Will Spijkers, *Greifbewegungen im virtuellen Raum - Auswirkungen von Pertubationen*, 53. Frühjahreskongress der GfA: Kompetenzentwicklung in realen und virtuellen Arbeitssystemen (Magdeburg, Germany) (GfA, ed.), February 28 - March 2, 2007, pp. 307–310.

[AWK+07b] _____, *Prehension in virtual environments: General aspects of grasping without haptic feedback and the effects of perturbations*, CyberTherapy '07: Proceedings of the 12th Annual CyberTherapy 2007 Conference: Transforming Healthcare Through Technology (Washington, DC), June 11-14, 2007.

[AWKS07] Solveyg Anders, Marc Wolter, Torsten Kuhlen, and Walter Sturm, *Three-dimensional uni- and crossmodal cueing in spatial shift of attention: a VR study*, CyberTherapy '07: Proceedings of the 12th Annual CyberTherapy 2007 Conference: Transforming Healthcare Through Technology (Washington, DC), June 11-14, 2007.

[Bar97] Thomas H. Barr, *Vector calculus*, Prentice-Hall, Upper Saddle River, NJ, 1997.

[BB98a] Paolo Baerlocher and Ronan Boulic, *Task priority formulations for the kinematic control of highly redundant articulated structures*, IROS '98: Proceedings of the IEEE Conference on Intelligent Robots and Systems (Victoria, Canada), 1998, pp. 323–329.

[BB98b] Rama Bindiganavale and Norman I. Badler, *Motion abstraction and mapping with spatial constraints*, CAPTECH'98: Proceedings of the International Workshop on Modelling and Motion Capture Techniques for Virtual Environments (Geneva, Switzerland) (Nadia Magnenat-Thalmann and Daniel Thalmann, eds.), Lecture Notes in Computer Science, vol. 1537/1998, Springer-Verlag Berlin / Heidelberg, November 26-28, 1998, pp. 70–82.

[BB04] Paolo Baerlocher and Ronan Boulic, *An inverse kinematic architecture enforcing an arbitrary number of strict priority levels*, The Visual Computer **20** (2004), no. 6, 402–417.

[BBR+07] Christian H. Bischof, H. Martin Bücker, Arno Rasch, Emil Slusanschi, and Bruno Lang, *Automatic differentiation of the general-purpose computational fluid dynamics package FLUENT*, Journal of Fluids Engineering **129** (2007), no. 5, 652–658.

[BCH+95] Ronan Boulic, Tolga K. Capin, Z. Huang, Prem Kalra, B. Linterrnann, Nadia Magnenat-Thalmann, Laurent Moccozet, Tom Molet, I. S. Pandzic, Kurt Saar, Alfred Schmitt, J. Shen, and Daniel Thalmann, *The HUMANOID environment for interactive animation of multiple deformable human characters*, Computer Graphics Forum **14** (1995), no. 3, 337–348.

[BCKM96] Christian H. Bischof, Alan Carle, Peyvand Khademi, and Andrew Mauer, *ADIFOR 2.0: Automatic differentiation of Fortran 77 programs*, IEEE Computational Science & Engineering **3** (1996), no. 3, 18–32.

[BCW+82] Roy A. Brand, Richard D. Crowninshield, C. E. Wittstock, Doug R. Pederson, Charles R. Clark, and F. M. Van Krieken, *A model of lower extremity muscular anatomy*, Journal of Biomechanical Engineering **104** (1982), no. 4, 304–310.

[BEL02] Norman I. Badler, Charles A. Erignac, and Ying Liu, *Virtual humans for validating maintenance procedures*, Communications of the ACM **45** (2002), no. 7, 56–63.

[Ber02] Rolf Bertolini, *Systematische Anatomie des Menschen*, Urban & Fischer Verlag, 2002.

[BFG+01] Alexandra Battaglia-Mayer, Stefano Ferraina, Aldo Genovesio, Barbara Marconi, Salvatore Squatrito, Marco Molinari, Francesco Lacquaniti, and Roberto Caminiti, *Eye–hand coordination during reaching. II. an analysis of the relationships between visuomanual signals in parietal cortex and parieto-frontal association projections*, Cerebral Cortex **11** (2001), no. 6, 528–544.

[BFM+98] Alexandra Battaglia-Mayer, Stefano Ferraina, Barbara Marconi, James B. Bullis, Francesco Lacquaniti, Yves Burnod, Pierre Baraduc, and R. Caminiti, *Early motor influences on visuomotor transformations for reaching: a positive image of optic ataxia*, Experimental Brain Research **123** (1998), no. 1–2, 172–189.

[BFM+00] Alexandra Battaglia-Mayer, Stefano Ferraina, Takashi Mitsuda, Barbara Marconi, Aldo Genovesio, Paolo Onorati, Francesco Lacquaniti, and Roberto Caminiti, *Early coding of reaching in the parietooccipital cortex*, Journal of Neurophysiology **83** (2000), no. 4, 2374–2391.

[BG95] Bruce M. Blumberg and Tinsley A. Galyean, *Multi-level direction of autonomous creatures for real-time virtual environments*, Proceedings of SIGGRAPH '95, 1995, pp. 47–54.

[BGBC05] Harry Brenton, Marco Gillies, Daniel Ballin, and David Chatting, *The uncanny valley: Does it exist?*, Proceedings of the Human-Animated Characters Interaction Workshop (as part of HCI 2005: The 19th British HCI Group Annual Conference) (Edinburgh, UK), September 6, 2005.

[BH00] Matthew Brand and Aaron Hertzmann, *Style machines*, Proceedings of SIGGRAPH '00, 2000, pp. 183–192.

[BHG93] Norman I. Badler, Michael J. Hollick, and John P. Granieri, *Real-time control of a virtual human using minimal sensors*, Presence: Teleoperators & Virtual Environments **2** (1993), no. 2, 82–86.

[Bio97] Frank Biocca, *The cyborg's dilemma: Embodiment in virtual environments*, Journal of Computer-Mediated Communication **3** (1997), no. 2, online journal.

[BKL+02] Nicole Beringer, Ute Kartal, Katerina Louka, Florian Schiel, and Uli Türk, *PROMISE: A procedure for multimodal interactive system evaluation*, Technical report, LMU München, 2002.

[BKLP05] Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola, and Ivan Poupyrev, *3D user interfaces: Theory and practice*, Addison-Wesley, 2005.

[Bli82] James F. Blinn, *A generalization of algebraic surface drawing*, ACM Transactions on Graphics **1** (1982), no. 3, 235–256.

[BMRS96] Frank Buschmann, Regine Meunier, Hans Rohnert, and Peter Sommerlad, *Pattern-oriented software architecture. A system of patterns*, John Wiley & Sons, 1996.

[BMS03] P. Paolo Battaglini, Amir Muzur, and Miran Skrap, *Visuomotor deficits and fast recovery after area V6A lesion in monkeys*, Behavioural Brain Research **139** (2003), no. 1–2, 115–122.

[BNB05] Volkert Buchmann, Trond Nilsen, and Mark Billinghurst, *Interaction with partially transparent hands and objects*, CRPIT '40: Proceedings of the Sixth Australasian Conference on User Interface (Darlinghurst, Australia), Australian Computer Society, Inc., 2005, pp. 17–20.

[BPB99] Norman I. Badler, Martha S. Palmer, and Rama Bindiganavale, *Animation control for real-time virtual humans*, Communications of the ACM **42** (1999), no. 8, 64–73.

[BPW93] Norman I. Badler, Cary B. Phillips, and Bonnie Lynn Webber, *Simulating humans: Computer graphics animation and control*, Oxford University Press, 1993.

[BRM97] Christian H. Bischof, Lucas Roh, and Andrew Mauer, *ADIC — An extensible automatic differentiation tool for ANSI-C*, Software: Practice and Experience **27** (1997), no. 12, 1427–1456.

[BRRP97] Bobby Bodenheimer, Chuck Rose, Seth Rosenthal, and John Pella, *The process of motion capture: Dealing with the data*, Proceedings of the Eurographics Computer Animation and Simulation Workshop (Wien, Austria), September 1997.

[Brü96] Godehard Brüntrup, *Das Leib-Seele-Problem. Eine Einführung*, Kohlhammer, 1996.

[BVW+07] Lydia Beck, René Vohn, Marc Wolter, Torsten Kuhlen, and Walter Sturm, *Spatial processing in near and far space using virtual reality methods: an*

fMRI-study, CyberTherapy '07: Proceedings of the 12th Annual CyberTherapy 2007 Conference: Transforming Healthcare Through Technology (Washington, DC), June 11-14, 2007.

[BW94] John H. Bailey and Bob G. Witmer, *Learning and transfer of spatial knowledge in a virtual environment*, Proceedings of the Human Factors and Ergonomics Society, 38th Annual Meeting (Santa Monica, CA), Human Factors and Ergonomics Society, 1994, pp. 1158–1162.

[BW95] Armin Bruderlin and Lance Williams, *Motion signal processing*, Proceedings of SIGGRAPH '95, 1995, pp. 97–104.

[CCSS91] Pasquale Chiacchio, Stefano Chiaverini, Lorenzo Sciavicco, and Bruno Siciliano, *Closed-loop inverse kinematics schemes for constrained redundant manipulators with task space augmentation and task priority strategy*, International Journal of Robotics Research **10** (1991), no. 4, 410–425.

[CGC+02] Steve Capell, Seth Green, Brian Curless, Tom Duchamp, and Zoran Popović, *Interactive skeleton-driven dynamic deformations*, Proceedings of SIGGRAPH '02, 2002, pp. 586–593.

[CGW+02] Heidi Chapman, Maria Gavrilescu, Hong Wang, Michael Kean, Gary Egan, and Umberto Castiello, *Posterior parietal cortex control of reach-to-grasp movements in humans*, European Journal of Neuroscience **15** (2002), no. 12, 2037–2042.

[CH05] Jinxiang Chai and Jessica K. Hodgins, *Performance animation from low-dimensional control signals*, Proceedings of SIGGRAPH '05, 2005, pp. 686–696.

[CHP89] John E. Chadwick, David R. Haumann, and Richard E. Parent, *Layered construction for deformable animated characters*, Proceedings of SIGGRAPH '89, 1989, pp. 243–252.

[CK00] Kwang-Jin Choi and Hyeong-Seok Ko, *On-line motion retargetting*, The Journal of Visualization and Computer Animation **11** (2000), no. 5, 223–235, Special Issue: Pacific Graphics '99.

[CPF05] Pietro Cerveri, Antonio Pedotti, and Giancarlo Ferrigno, *Kinematical models to reduce the effect of skin artifacts on marker-based human motion estimation*, Journal of Biomechanics **38** (2005), no. 11, 2228–2236.

[CRPF03] Pietro Cerveri, Marco Rabuffetti, Antonio Pedotti, and Giancarlo Ferrigno, *Real-time human motion estimation using biomechanical models and non-linear state-space filters*, Medical & Biological Engineering & Computing **41** (2003), no. 2, 109–123.

[CVB01] Justine Cassell, Hannes Högni Vilhjálmsson, and Timothy Bickmore, *BEAT: the behavior expression animation toolkit*, Proceedings of SIGGRAPH '01, 2001, pp. 477–486.

[CZ92] David T. Chen and David Zeltzer, *Pump it up: computer animation of a biomechanically based model of muscle using the finite element method*, Proceedings of SIGGRAPH '92, 1992, pp. 89–98.

[DBC+06] Nicole David, Bettina H. Bewernick, Michael X Cohen, Albert Newen, Silke Lux, Gereon R. Fink, N. Jon Shah, and Kai Vogeley, *Neural representations of self versus other: Visual–spatial perspective taking and agency in a virtual ball-tossing game*, Journal of Cognitive Neuroscience **18** (2006), no. 6, 898–910.

[DCFN06] Zhigang Deng, Pei-Ying Chiang, Pamela Fox, and Ulrich Neumann, *Animating blendshape faces by cross-mapping motion capture data*, SI3D '06: Proceedings of the 2006 symposium on Interactive 3D graphics and games, 2006, pp. 43–48.

[DCKY02] Feng Dong, Gordon J. Clapworthy, Meleagros A. Krokos, and Jialian Yao, *An anatomy-based approach to human muscle modeling and deformation*, IEEE Transactions on Visualization and Computer Graphics **8** (2002), no. 2, 154–170.

[DFM05] Paula J. Durlach, Jennifer Fowlkes, and Christopher J. Metevier, *Effect of variations in sensory feedback on performance in a virtual reaching task*, Presence: Teleoperators & Virtual Environments **14** (2005), no. 4, 450–462.

[DH55] J. Denavit and R. S. Hartenberg, *A kinematic notation for lower- pair mechanism based on matrices*, Journal of Applied Mechanics **22** (1955), no. 2, 215–221.

[dJvdGP01] B. M. de Jong, F. H. C. E. van der Graaf, and A. M. J. Paans, *Brain activation related to the representations of external space and body scheme in visuomotor control*, NeuroImage **14** (2001), no. 5, 1128–1135.

[DKL98] Erik B. Dam, Martin Koch, and Martin Lillholm, *Quaternions, interpolation and animation*, Technical Report DIKU-TR-98/5, Department of Computer Science, University of Copenhagen, 1998.

[DKSF04] Christian Dohle, Raimund Kleiser, Rüdiger J. Seitz, and Hans-Joachim Freund, *Body scheme gates visual processing*, Journal of Neurophysiology **91** (2004), no. 5, 2376–2379.

[DL95] Scott L. Delp and J. Peter Loan, *A graphics-based software system to develop and analyze models of musculoskeletal structures*, Computers in Biology and Medicine **25** (1995), no. 1, 21–34.

[Dod04] Neil A. Dodgson, *Variation and extrema of human interpupillary distance*, SPIE '04: Proceedings of Stereoscopic Displays and Virtual Reality Systems (San Jose, CA), January 19-22, 2004, pp. 36–46.

[DRBZ01] Craig J. Della Valle, Andrew S. Rokito, Maureen Gallagher Birdzell, and Joseph D. Zuckerman, *Biomechanics of the shoulder*, Basic Biomechanics of

the Musculoskeletal System (Margareta Nordin and Victor H. Frankel, eds.), Lippincott Williams & Wilkins, Baltimore, MD, 3rd ed., 2001, pp. 318–339.

[Dut97] Thierry Dutoit, *High-quality text-to-speech synthesis: an overview*, Journal of Electrical & Electronics Engineering **17** (1997), 25–37.

[EAPL06] Claudia Esteves, Gustavo Arechavaleta, Julien Pettré, and Jean-Paul Laumond, *Animation planning for virtual characters cooperation*, ACM Transactions on Graphics **25** (2006), no. 2, 319–339.

[EDM+97] Stephen R. Ellis, Nancy. S. Dorighi, Brian. M. Menges, Bernard D. Adelstein, and Richard H. Jacoby, *In search of equivalence classes in subjective scales of reality*, HCI '97: Proceedings of the 7th International Conference on Human Computer Interaction (San Francisco, CA), August 24-29, 1997, pp. 873–876.

[FC68] Seymour Fisher and Sidney E. Cleveland, *Body image and personality*, Dover Publications, New York, 1968.

[Fer03] Aaron P. Ferguson, *Skin deep beauty: A production friendly creature geometry pipeline used on hulk*, SIGGRAPH 2003: Sketches & Applications, 2003.

[FFG+03] C. Farrer, N. Franck, N. Georgieff, C. D. Frith, J. Decety, and M. Jeannerod, *Modulating the experience of agency: a positron emission tomography study*, NeuroImage **18** (2003), no. 2, 324–333.

[FGKG01] Patrizia Fattori, Michela Gamberini, Dieter F. Kutz, and Claudio Galletti, *'Arm-reaching' neurons in the parietal area V6A of the macaque monkey*, European Journal of Neuroscience **13** (2001), no. 12, 2309–2313.

[FL01] J. Randall Flanagan and Sarah Lolley, *The inertial anisotropy of the arm is accurately predicted during movement planning*, The Journal of Neuroscience **21** (2001), no. 4, 1361–1369.

[FP85] Giancarlo Ferrigno and Antonio Pedotti, *ELITE: A digital dedicate hardware system for movement analysis via real-time TV signal processing*, IEEE Transactions on Biomedical Engineering **32** (1985), no. 11, 943–950.

[FTT99] John Funge, Xiaoyuan Tu, and Demetri Terzopoulos, *Cognitive modeling: knowledge, reasoning and planning for intelligent characters*, Proceedings of SIGGRAPH '99, 1999, pp. 29–38.

[Ger04] Margaret S. Geroch, *Motion capture for the rest of us*, Journal of Computing Sciences in Colleges **19** (2004), no. 3, 157–164.

[GFGK99] Claudio Galletti, Patrizia Fattori, Michela Gamberini, and Dieter F. Kutz, *The cortical visual area V6: brain location and visual topography*, European Journal of Neuroscience **11** (1999), no. 11, 3922–3936.

[GFKB97] Claudio Galletti, Patrizia Fattori, Dieter F. Kutz, and P. P. Battaglini, *Arm movement-related neurons in the visual area V6A of the macaque superior parietal lobule*, European Journal of Neuroscience **9** (1997), no. 2, 410–413.

[GFKG99] Claudio Galletti, Patrizia Fattori, Dieter F. Kutz, and Michela Gamberini, *Brain location and visual topography of cortical area V6A in the macaque monkey*, European Journal of Neuroscience **11** (1999), no. 2, 575–582.

[GGK+01] Claudio Galletti, Michela Gamberini, Dieter F. Kutz, Patrizia Fattori, Giuseppe Luppino, and Massimo Matelli, *The cortical connections of area v6: an occipito-parietal network processing visual information*, European Journal of Neuroscience **13** (2001), no. 8, 1572–1588.

[Gle98] Michael Gleicher, *Retargetting motion to new characters*, Proceedings of SIGGRAPH '98, 1998, pp. 33–42.

[Gle01a] ———, *Motion path editing*, SI3D '01: Proceedings of the 2001 symposium on Interactive 3D graphics (Research Triangle Park, NC), March 19-21, 2001, pp. 195–202.

[Gle01b] Micheal Gleicher, *Comparing constraint-based motion editing methods*, Graphical Models **63** (2001), no. 2, 107–134.

[GMCW05] Yan Gao, Lizhuang Ma, Zhihua Chen, and Xiaomao Wu, *Motion normalization: the preprocess of motion data*, VRST '05: Proceedings of the ACM symposium on Virtual Reality Software and Technology (Monterey, CA), 2005, pp. 253–256.

[GMHP04] Keith Grochow, Steven L. Martin, Aaron Hertzmann, and Zoran Popović, *Style-based inverse kinematics*, Proceedings of SIGGRAPH '04, 2004, pp. 522–531.

[GMT89] Jean-Paul Gourret, Nadia Magnenat-Thalmann, and Daniel Thalmann, *Simulation of object and human skin formations in a grasping task*, Proceedings of SIGGRAPH '89, 1989, pp. 21–30.

[GNV+04] Martina Graf, H.-C. Nuerk, Jakob T. Valvoda, Torsten Kuhlen, and Klaus Willmes, *Wie sind Zahlen mental im 3D-Raum repräsentiert? [how are numbers represented in the 3d space?]*, Proceedings of the 44th DGP Congress, 2004.

[Gra98] F. Sebastian Grassia, *Practical parameterization of rotations using the exponential map*, Journal of Graphics Tool **3** (1998), no. 3, 29–48.

[Gra05] Henry Gray, *Gray's anatomy: the anatomical basis of clinical practice.*, 39th ed., Elsevier Churchill Livingstone, Edinburgh, New York, 2005.

[Gri00] Andreas Griewank, *Evaluating derivatives: Principles and techniques of algorithmic differentiation*, Frontiers in Applied Mathematics, no. 19, SIAM, Philadelphia, PA, 2000.

[GSI+04] Stefan Göbel, Oliver Schneider, Ido Iurgel, Axel Feix, Christian Knöpfle, and Alexander Rettig, *Virtual Human: Storytelling and computer graphics for a virtual human platform*, Proceedings of the 2nd International Conference on

Technologies for Interactive Digital Storytelling and Entertainment (Darmstadt, Germany) (Stefan Göbel and Ulrike Spierling, eds.), Lecture Notes in Computer Science, vol. 3105, Springer-Verlag, June 24-26, 2004, pp. 79–88.

[GSKJ03] Michael Gleicher, Hyun Joon Shin, Lucas Kovar, and Andrew Jepsen, *Snap-together motion: assembling run-time animations*, SI3D '03: Proceedings of the 2003 symposium on Interactive 3D graphics (Monterey, CA), 2003, pp. 181–188.

[GvWP04] Ken W. Grant, Virginie van Wassenhove, and David Poeppel, *Detection of auditory (cross-spectral) and auditory-visual (cross-modal) synchrony*, Speech Communication **44** (2004), 43–53.

[GW05] Zheng Guo and Kok Cheong Wong, *Skinning with deformable chunks*, Computer Graphics Forum **24** (2005), no. 3, 373–382.

[GWN+03] Markus Gross, Stephan Würmlin, Martin Naef, Edouard Lamboray, Christian Spagno, Andreas Kunz, Esther Koller-Meier, Tomas Svoboda, Luc Van Gool, Silke Lang, Kai Strehlke, Andrew Vande Moere, and Oliver Staadt, *blue-c: A spatially immersive display and 3d video portal for telepresence*, ACM Transactions on Graphics **22** (2003), no. 3, 819–827.

[H-A04] H-Anim Working Group, *H-Anim - humanoid animation 200x*, 2004, International Standard ISO/IEC FCD 19774:200x.

[Har78] David E. Hardt, *Determining muscle forces in the leg during normal human walking – an application and evaluation of optimization methods*, Journal of Biomechanical Engineering **100** (1978), no. 2, 72–78.

[Hat76] Herbert Hatze, *The complete optimization of a human motion*, Mathematical Bioscience **28** (1976), no. 1–2, 99–135.

[HBB04] Vassili Hurmusiadis, Simon Barrick, and Chris Briscoe, *Visualization of muscle function for medical education*, MMVR '04: Proceedings of Medicine Meets Virtual Reality 12 (Newport Beach, CA) (J.D. Westwood et al., ed.), IOS Press, Amsterdam, Jan 15-17, 2004, pp. 137–143.

[Hee92] Carrie Heeter, *Being there: the subjective experience of presence*, Presence: Teleoperators & Virtual Environments **1** (1992), no. 2, 262–271.

[Hei50] Robert A. Heinlein, *Waldo & Magic Inc.*, Del Rey, New York, 1950.

[HFP+00] Lorna Herda, Pascal Fua, Ralf Plänkers, Ronan Boulic, and Daniel Thalmann, *Skeleton-based motion capture for robust reconstruction of human motion*, CA '00: Proceedings of Computer Animation (Philadelphia, PA), May 3-5, 2000, pp. 77–85.

[HH93] Deborah Hix and H. Rex Hartson, *Developing user interfaces: Ensuring usability through product & process*, John Wiley, New York, 1993.

[Hil38] Archibald V. Hill, *The heat of shortening and the dynamic constants of muscle*, Proceedings of the Royal Society of London, Series B: Biological Sciences **126** (1938), 136–195.

[HJCW06] Min Hong, Sunhwa Jung, Min-Hyung Choi, and S. W. J. Welch, *Fast volume preservation for a mass-spring system*, IEEE Computer Graphics and Applications **26** (2006), no. 5, 83–91.

[HP00] Jessica K. Hodgins and Zoran Popović, *Animating humans by combining simulation and motion capture*, SIGGRAPH '00: Course Notes, 2000.

[HPP05] Eugene Hsu, Kari Pulli, and Jovan Popović, *Style translation for human motion*, Proceedings of SIGGRAPH '05, 2005, pp. 1082–1089.

[HSK05] Alexander Hornung, Sandip Sar-Dessai, and Leif Kobbelt, *Self-calibrating optical motion tracking for articulated bodies*, VR '05: Proceedings of the IEEE Virtual Reality Conference (Bonn, Germany), March 12-16, 2005, pp. 75–82.

[HVK+04] Ines A. Heber, Jakob T. Valvoda, Torsten Kuhlen, Walter Sturm, and Bruno Fimm, *Visuelle Aufmerksamkeitsausrichtung im virtuellen Raum*, Zeitschrift für Neuropsychologie **15** (2004), 121.

[HVK+05] ———, *Asymmetries of visuo-spatial attention in virtual space after sleep deprivation*, Journal of Cognitive Neuroscience **Suppl. S** (2005), 51.

[HWKF07] Ines Ann Heber, Marc Wolter, Torsten Kuhlen, and Bruno Fimm, *Asymmetries of covert attention in virtual space*, CyberTherapy '07: Proceedings of the 12th Annual CyberTherapy 2007 Conference: Transforming Healthcare Through Technology (Washington, DC), June 11-14, 2007.

[HYC+05] Dae-Eun Hyun, Seung-Hyun Yoon, Jung-Woo Chang, Joon-Kyung Seong, Myung-Soo Kim, and Bert Jüttler, *Sweep-based human deformation*, The Visual Computer **21** (2005), no. 8–10, 542–550.

[IC05] Lucio Ieronutti and Luca Chittaro, *A virtual human architecture that integrates kinematic, physical and behavioral aspects to control H-Anim characters*, Proceedings of the 10th International Conference on 3D Web Technology (Bangor, United Kingdom), March 29 - April 1, 2005.

[IdRFA00] Wijnand A. IJsselsteijn, Huib de Ridder, Jonathan Freeman, and S. E. Avons, *Presence: Concept, determinants and measurement*, SPIE '00: Proceedings of Stereoscopic Displays and Virtual Reality Systems (San Jose, CA), January 23-28, 2000, pp. 3959–3976.

[Ish04] Hiroshi Ishii, *Bottles: A transparent interface as a tribute to Mark Weiser*, IEICE Transactions on Information and Systems **E87-D** (2004), no. 6, 1299–1311.

[Jac01] David Jacobson, *Presence revisited: Imagination, competence, and activity in text-based virtual worlds*, CyberPsychology & Behavior **4** (2001), no. 6, 653–673.

[JD75] Ronald H. Jensen and Dwight T. Davy, *An investigation of muscle lines of actions about the hip: A centroid approach vs. the straight line approach*, Journal of Biomechanics **8** (1975), no. 2, 103–110.

[JR97] W. Lewis Johnson and Jeff Rickel, *Steve: an animated pedagogical agent for procedural training in virtual environments*, SIGART Bulletin **8** (1997), no. 1-4, 16–21.

[JTDP03] Pushkar Joshi, Wen C. Tien, Mathieu Desbrun, and Frédéric Pighin, *Learning controls for blend shape based realistic facial animation*, SCA '03: Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (Aire-la-Ville, Switzerland), Eurographics Association, 2003, pp. 187–192.

[JWL05] Pieter Jorissen, Maarten Wijnants, and Wim Lamotte, *Dynamic interactions in physically realistic collaborative virtual environments*, IEEE Transactions on Visualization and Computer Graphics **11** (2005), no. 6, 649–660.

[KG03] Lucas Kovar and Michael Gleicher, *Flexible automatic motion blending with registration curves*, SCA '03: Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (San Diego, CA), Eurographics Association, 2003, pp. 214–224.

[KGP02] Lucas Kovar, Michael Gleicher, and Frédéric Pighin, *Motion graphs*, Proceedings of SIGGRAPH '02, 2002, pp. 473–482.

[KHS03] Kolja Kähler, Jörg Haber, and Hans-Peter Seidel, *Reanimating the dead: reconstruction of expressive faces from skull data*, Proceedings of SIGGRAPH '03, 2003, pp. 554–561.

[KJP02] Paul G. Kry, Doug L. James, and Dinesh K. Pai, *Eigenskin: real time large deformation character skinning in hardware*, SCA '02: Proceedings of the 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, 2002, pp. 153–159.

[KKKL98] G. Jounghyun Kim, Kyo Chul Kang, Hyejung Kim, and Jiyoung Lee, *Software engineering of virtual worlds*, VRST '98: Proceedings of the ACM symposium on Virtual Reality Software and Technology, 1998, pp. 131–138.

[KKS95] Myoung-Jun Kim, Myung-Soo Kim, and Sung Yong Shin, *A general construction scheme for unit quaternion curves with simple high order derivatives*, Proceedings of SIGGRAPH '95, 1995, pp. 369–376.

[KKS00] Torsten Kuhlen, Karl-Friedrich Kraiss, and Roland Steffan, *How VR-based reach-to-grasp experiments can help to understand movement organization within the human brain*, Presence: Teleoperators & Virtual Environments **9** (2000), no. 4, 350–359.

[KLBL93] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal, *Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness*, International Journal of Aviation Psychology **3** (1993), no. 3, 203–220.

[Kle89] J. Kleiser, *A fast, efficient, accurate way to represent the human face*, SIGGRAPH '89: Course Notes, 1989.

[KMM+98] Prem Kalra, Nadia Magnenat-Thalmann, Laurent Moccozet, Gael Sannier, Amaury Aubel, and Daniel Thalmann, *Real-time animation of realistic virtual humans*, IEEE Computer Graphics and Applications **18** (1998), no. 5, 42–56.

[KSG02] Lucas Kovar, John Schreiner, and Michael Gleicher, *Footskate cleanup for motion capture editing*, SCA '02: Proceedings of the 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (San Antonio, TX), July 21-22, 2002, pp. 97–104.

[KSK00] Taku Komura, Yoshihisa Shinagawa, and Tosiyasu L. Kunii, *Creating and retargetting motion by the musculoskeletal human body model*, The Visual Computer **16** (2000), no. 5, 254–270.

[KSK01] ———, *An inverse kinematics method based on muscle dynamics*, CGI '01: Proceedings of Computer Graphics International, 2001, pp. 15–22.

[KW04] Stefan Kopp and Ipke Wachsmuth, *Synthesizing multimodal utterances for conversational agents*, Computer Animation and Virtual Worlds **15** (2004), no. 1, 39–52.

[Lan98a] Jeff Lander, *Making kine more flexible*, Game Developer Magazine **5** (1998), no. 11, 15–22.

[Lan98b] ———, *Skin them bones: Game programming for the web generation*, Game Developer Magazine **5** (1998), no. 5, 11–16.

[LCF00] J. P. Lewis, Matt Cordner, and Nickson Fong, *Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation*, Proceedings of SIGGRAPH '00, 2000, pp. 165–172.

[LCL06] Kang Hoon Lee, Myung Geol Choi, and Jehee Lee, *Motion patches: building blocks for virtual environments annotated with motion data*, ACM Transactions on Graphics **25** (2006), no. 3, 898–906.

[LCR+02] Jehee Lee, Jinxiang Chai, Paul S. A. Reitsma, Jessica K. Hodgins, and Nancy S. Pollard, *Interactive control of avatars animated with human motion data*, Proceedings of SIGGRAPH '02, 2002, pp. 491–500.

[LD97] Matthew Lombard and Theresa Ditton, *At the heart of it all: The concept of presence*, Journal of Computer-Mediated Communication **3** (1997), no. 2, online journal.

[LDA+02] James Jeng-Weei Lin, Henry B. L. Duh, Habib Abi-Rached, Donald E. Parker, and Thomas A. Furness Iii, *Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment*, VR '02: Proceedings of the IEEE Virtual Reality Conference (Washington, DC), March 24-28, 2002, pp. 164–171.

[LKK00] Ji Y. Lee, Hye J. Kim, and Kyo C. Kang, *A real world object modeling method for creating simulation environment of real-time systems*, OOPSLA '00: Proceedings of the 15th ACM SIGPLAN Conference on Object-oriented Programming, Systems, Languages, and Applications (Minneapolis, MN), 2000, pp. 93–104.

[LMDN05] J. P. Lewis, Jonathan Mooser, Zhigang Deng, and Ulrich Neumann, *Reducing blendshape interference by selected motion attenuation*, SI3D '05: Proceedings of the 2005 symposium on Interactive 3D graphics and games (New York, NY), 2005, pp. 25–29.

[Loo92] Jack M. Loomis, *Distal attribution and presence*, Presence: Teleoperators & Virtual Environments **1** (1992), no. 1, 113–118.

[LP02] C. Karen Liu and Zoran Popović, *Synthesis of complex dynamic character motion from simple animations*, Proceedings of SIGGRAPH '02, 2002, pp. 408–416.

[LS99] Jehee Lee and Sung Yong Shin, *A hierarchical approach to interactive motion editing for human-like figures*, Proceedings of SIGGRAPH '99, 1999, pp. 39–48.

[LSVA07] Tobias Lentz, Dirk Schröder, Michael Vorländer, and Ingo Assenmacher, *Virtual reality system with integrated sound field simulation and reproduction*, EURASIP Journal on Advances in Signal Processing **2007** (2007), 19.

[LT06] Sung-Hee Lee and Demetri Terzopoulos, *Heads up!: biomechanical modeling and neuromuscular control of the neck*, ACM Transactions on Graphics **25** (2006), no. 3, 1188–1198.

[LTW95] Yuencheng Lee, Demetri Terzopoulos, and Keith Walters, *Realistic modeling for facial animation*, Proceedings of SIGGRAPH '95, 1995, pp. 55–62.

[Lue03] David G. Luenberger, *Linear and nonlinear programming*, 2nd ed., Kluwer Academic, Boston, 2003.

[LWS02] Yan Li, Tianshu Wang, and Heung-Yeung Shum, *Motion texture: a two-level statistical model for character motion synthesis*, Proceedings of SIGGRAPH '02, 2002, pp. 465–472.

[LZWM06] Guodong Liu, Jingdan Zhang, Wei Wang, and Leonard McMillan, *Human motion estimation from a reduced marker set*, SI3D '06: Proceedings of the 2006 symposium on Interactive 3D graphics and games (Redwood City, CA), 2006, pp. 35–42.

[Mac90] Anthony A. Maciejewski, *Dealing with the ill-conditioned equations of motion for articulated figures*, IEEE Computer Graphics and Applications **10** (1990), no. 3, 63–71.

[MAN97] James Maida, A. Aldridge, and Jennifer Novak, *Effects of lighting on human performance in training.*, HCI '97: Proceedings of the 7th International Conference on Human Computer Interaction (San Francisco, CA), August 24-29, 1997, pp. 877–880.

[MBBT00] Jean-Sébastien Monzani, Paolo Baerlocher, Ronan Boulic, and Daniel Thalmann, *Using an intermediate skeleton and inverse kinematics for motion retargeting*, Proceedings of Eurographics 2000 (Interlaken, Switzerland), August 2000.

[MBC01] Mark Mizuguchi, John Buchanan, and Tom Calvert, *Data driven motion transitions for interactive games*, Eurographics 2001: Short Presentations (Manchester, United Kingdom), September 2001.

[MBT99] Tom Molet, Ronan Boulic, and Daniel Thalmann, *Human motion capture driven by orientation measurements*, Presence: Teleoperators & Virtual Environments **8** (1999), no. 2, 187–203.

[MC94] Dinesh Manocha and John F. Canny, *Efficient inverse kinematics for general 6r manipulators*, IEEE Transactions on Robotics and Automation **10** (1994), no. 5, 648–657.

[MG03] Alex Mohr and Michael Gleicher, *Building efficient, accurate character skins from examples*, ACM Transactions on Graphics **22** (2003), no. 3, 562–568.

[Mil00] Jan-Torsten Milde, *Lokutor: Towards a believable communicative agent*, AGENTS 2000: Proceedings of the Fourth International Conference on Autonomous Agents, Achieving Human-Like Behavior in Interactive Animated Agents (Barcelona, Spain) (Jeff Rickel, W. Lewis Johnson, and James Lester, eds.), June 3-7, 2000.

[Min80] Marvin Minsky, *Telepresences*, Omni (1980), 45–51.

[MIWB01] Michael Meehan, Brent Insko, Mary Whitton, and Frederick P. Brooks, *Physiological measures of presence in virtual environments*, PRESENCE 2001: Proceedings of the 4th Annual International Workshop (Philadelphia, PA), May 21-23, 2001.

[MIWB02] ———, *Physiological measures of presence in stressful virtual environments*, Proceedings of SIGGRAPH '02, 2002, pp. 645–652.

[MKKP01] Franz Mechsner, Dirk Kerzel, Günther Knoblich, and Wolfgang Prinz, *Perceptual basis of bimanual coordination*, Nature **414** (2001), no. 6859, 69–73.

[MM04a] Andrea H. Mason and Christine L. MacKenzie, *The role of graphical feedback about self-movement when receiving objects in an augmented environment*, Presence: Teleoperators & Virtual Environments **13** (2004), no. 5, 507–519.

[MM04b] Michael Meredith and Steve Maddock, *Real-time inverse kinematics: The return of the jacobian*, Technical Report CS-04-06, Department of Computer Science, University of Sheffield, 2004.

[MMG06] Bruce Merry, Patrick Marais, and James Gain, *Animation space: A truly linear framework for character animation*, ACM Transactions on Graphics **25** (2006), no. 4, 1400–1423.

[MOCC05] Donald W. Massaro, Slim Ouni, Michael M. Cohen, and Rashid Clark, *A multilingual embodied conversational agent*, Proceedings of the HICCS'05 conference (Los Alimitos, CA), 2005, p. 296.2.

[Mor70] Masahiro Mori, *Bukimi no tani [the uncanny valley]*, Energy **7** (1970), no. 4, 33–35, (in Japanese).

[MT96] Nadia Magnenat-Thalmann and Daniel Thalmann, *Computer animation*, ACM Computing Surveys **28** (1996), no. 1, 161–163.

[NB03] David Nunez and Edwin Blake, *A direct comparison of presence levels in text-based and graphics-based virtual environments*, AFRIGRAPH '03: Proceedings of the 2nd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa (Cape Town, South Africa), February 3-5, 2003, pp. 53–56.

[NF97] Victor Ng-Thow-Hing and Eugene Fiume, *Interactive display and animation of b-spline solids as muscle shape primitives*, Eurographics Workshop on Animation and Simulation 1997 (Budapest, Hungary), Springer-Verlag, Wien, September 1–2, 1997, pp. 81–97.

[NF02] ———, *Application-specific muscle representations*, Proceedings of Graphics Interface, 2002, pp. 107–116.

[NH86] Yoshihiko Nakamura and Hideo Hanafusa, *Inverse kinematic solutions with singularity robustness for robot manipulator control*, Journal of Dynamic Systems, Measurement, and Control **108** (1986), 163–171.

[NHK+85] Hitoshi Nishimura, Makoto Hirai, Toshiyuki Kawai, Toru Kawata, Isao Shirkawa, and Koichi Omura, *Object modeling by distribution function and a method of image generation*, Transactions of the Institute of Electronics and Communication Engineers of Japan **J68D** (1985), no. 4, 718–725, Proceedings of Electronics Communication Conference '85.

[NHY87] Yoshihiko Nakamura, Hideo Hanafusa, and Tsuneo Yoshikawa, *Task-priority based redundancy control of robot manipulators*, International Journal of Robotics Research **6** (1987), no. 2, 3–15.

[NT98] Luciana Porcher Nedel and Daniel Thalmann, *Modeling and deformation of the human body using an anatomically-based approach*, CA '98: Proceedings of the Computer Animation, 1998, p. 34.

[OBBH00] James F. O'Brien, Robert Bodenheimer, Gabriel J. Brostow, and Jessica K. Hodgins, *Automatic joint parameter estimation from magnetic motion capture data*, Proceedings of the Graphics Interface 2000 Conference (Montréal, Québec, Canada) (Sidney Fels and Pierre Poulin, eds.), Canadian Human-Computer Communications Society, May 15-17 2000, pp. 53–60.

[OC00] Sharon Oviatt and Philip Cohen, *Perceptual user interfaces: multimodal interfaces that process what comes naturally*, Communications of the ACM **43** (2000), no. 3, 45–53.

[Old71] R. C. Oldfield, *The assessment and analysis of handedness: The Edinburgh inventory*, Neuropsychologia **9** (1971), 97–113.

[OS84] David E. Orin and William W. Schrader, *Efficient computation of the Jacobian for robot manipulators*, International Journal of Robotics Research **3** (1984), no. 4, 66–75.

[Pag02] Horst Pagel, *Muskelphysiologie: Wie funktioniert das Schnitzel, bevor es auf dem Teller landet?*, Technical report, Institut für Physiologie, Universität zu Lübeck, 2002.

[Par72] Frederick I. Parke, *Computer generated animation of faces*, ACM '72: Proceedings of the ACM Annual Conference (New York, NY), 1972, pp. 451–457.

[PB88] Cary B. Phillips and Norman I. Badler, *JACK: a toolkit for manipulating articulated figures*, UIST '88: Proceedings of the 1st annual ACM SIGGRAPH Symposium on User Interface Software, 1988, pp. 221–229.

[PB02] Katherine Pullen and Christoph Bregler, *Motion capture assisted animation: texturing and synthesis*, Proceedings of SIGGRAPH '02, 2002, pp. 501–508.

[PCLS05] Michael Pratscher, Patrick Coleman, Joe Laszlo, and Karan Singh, *Outside-in anatomy based character rigging*, SCA '05: Proceedings of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (Los Angeles, CA), 2005, pp. 329–338.

[PCMH02] Kerstin D. Pfann, Daniel M. Corcos, Charity G. Moore, and Ziaul Hasan, *Circle-drawing movements at different speeds: Role of inertial anisotropy*, Journal of Neurophysiology **88** (2002), no. 5, 2399–2407.

[PFB+01] Daniela Perani, Ferruccio Fazio, Nunzio Alberto Borghese, Marco Tettamanti, Stefano Ferrari, Jean Decety, and Maria Carla Gilardi, *Different brain correlates for watching real and virtual hand actions*, NeuroImage **14** (2001), no. 3, 749–758.

[PG96] Ken Perlin and Athomas Goldberg, *Improv: a system for scripting interactive actors in virtual worlds*, Proceedings of SIGGRAPH '96, 1996, pp. 205–216.

[PH06] Sang Il Park and Jessica K. Hodgins, *Capturing and animating skin deformation in human motion*, ACM Transactions on Graphics **25** (2006), no. 3, 881–889.

[PHL+98] Frédéric Pighin, Jamie Hecker, Dani Lischinski, Richard Szeliski, and David H. Salesin, *Synthesizing realistic facial expressions from photographs*, Proceedings of SIGGRAPH '98, 1998, pp. 75–84.

[PI04] Helmut Prendinger and Mitsuru Ishizuka (eds.), *Life-like characters: Tools, affective functions, and applications*, Cognitive Technologies, Springer-Verlag, 2004.

[PL05] Norbert Pfleger and Markus Löckelt, *Synchronizing dialogue contributions of human users and virtual characters in a virtual reality environment*, Proceedings of Interspeech'2005 - Eurospeech 9th European Conference on Speech Communication and Technology (Lisbon, Portugal), 2005, pp. 2773–2776.

[PM02] Peter Poller and Jochen Müller, *Distributed audio-visual speech synchronization*, Proceedings of the ICSLP '02 Conference, 2002.

[PPFW95] Jerrold D. Prothero, Donald E. Parker, Thomas A. Furness, and Maxwell J. Wells, *Towards a robust, quantitative measure for presence.*, Proceedings of the Conference on Experimental Analysis and Measurement of Situation Awareness (Daytona Beach, FL), November 1-3, 1995, pp. 359–366.

[PPM+03] Michal Ponder, George Papagiannakis, Tom Molet, Nadia Magnenat-Thalmann, and Daniel Thalmann, *VHD++ development framework: Towards extendible, component based VR/AR simulation engine featuring advanced virtual character technologies*, CGI '03: Proceedings of Computer Graphics International, 2003, pp. 96–104.

[PPW97] Randy Pausch, Dennis Proffitt, and George Williams, *Quantifying immersion in virtual reality*, Proceedings of SIGGRAPH '97, 1997, pp. 13–18.

[Pri04] Primal Pictures, *Complete human anatomy: Primal 3d interactive series*, Primal Pictures Ltd., London, UK, 2004.

[PSM81] Richard P. Paul, Bruce Shimano, and Gordon E. Mayer, *Kinematic control equations for simple manipulators*, IEEE Transactions on Systems, Man, and Cybernetics **SMC-116** (1981), no. 6, 449–455.

[PSS02] Sang Il Park, Hyun Joon Shin, and Sung Yong Shin, *On-line locomotion generation based on motion blending*, SCA '02: Proceedings of the 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (San Antonio, TX), 2002, pp. 105–111.

[PT85] Jiapu Pan and Willis J. Tompkins, *A real-time QRS detection algorithm*, IEEE Transactions on Biomedical Engineering **BME-32** (1985), no. 3, 230–236.

[PV88] M.-T. Perenin and A. Vighetto, *Optic ataxia: a specific disruption in visuomotor mechanisms. I. different aspects of the deficit in reaching for objects*, Brain **111** (1988), no. 3, 643–674.

[PW99] Zoran Popović and Andrew Witkin, *Physically based motion transformation*, Proceedings of SIGGRAPH '99, 1999, pp. 11–20.

[RAB+03] Norbert Reithinger, Jan Alexandersson, Tilman Becker, Anselm Blocher, Ralf Engel, Markus Löckelt, Jochen Müller, Norbert Pfleger, Peter Poller, Michael Streit, and Valentin Tschernomas, *SmartKom - adaptive and flexible multimodal access to multiple applications*, Proceedings of ICMI 2003 Conference (Vancouver, B.C., Canada), 2003, pp. 101–108.

[Rai92] Rositsa T. Raikova, *A general approach for modelling and mathematical investigation of the human upper limb*, Journal of Biomechanics **25** (1992), no. 8, 857–867.

[RDI03] Giuseppe Riva, Fabrizio Davide, and Wijnand A. IJsselsteijn (eds.), *Being there: Concepts, effects and measurements of user presence in synthetic environments*, Emerging Communication: Studies on New Technologies and Practices in Communication, vol. 5, IOS Press, Amsterdam, 2003.

[Ree91] Byron Reeves, *'Being there': Television as symbolic versus natural experience.*, Unpublished manuscript, Stanford University, Institute for Communication Research, Stanford, CA, 1991.

[RGBC96] Charles Rose, Brian Guenter, Bobby Bodenheimer, and Michael F. Cohen, *Efficient generation of motion transitions using spacetime constraints*, Proceedings of SIGGRAPH '96, 1996, pp. 147–154.

[RJ99] Jeff Rickel and W. Lewis Johnson, *Animated agents for procedural training in virtual reality: Perception, cognition, and motor control*, Applied Artificial Intelligence **13** (1999), no. 4-5, 343–382.

[RM97] Giuseppe Riva and Luca Melis, *Virtual reality for the treatment of body image disturbances*, Virtual Reality in Neuro-Psycho-Physiology, 1997.

[RP04] Paul S. A. Reitsma and Nancy S. Pollard, *Evaluating motion graphs for character navigation*, SCA '04: Proceedings of the 2004 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (Grenoble, France), 2004, pp. 89–98.

[RS02] Holger Regenbrecht and Thomas Schubert, *Real and Illusory Interactions Enhance Presence in Virtual Environments*, Presence: Teleoperators & Virtual Environments **11** (2002), no. 4, 378–403.

[RSH+05] Liu Ren, Gregory Shakhnarovich, Jessica K. Hodgins, Hanspeter Pfister, and Paul Viola, *Learning silhouette features for control of human motion*, ACM Transactions on Graphics **24** (2005), no. 4, 1303–1331.

[SA89] Ali A. Seireg and R. Arvikar, *Biomechanical analysis of the musculoskeletal structure for medicine & sports*, Hemisphere Publishing Corporation, 1989.

[SA03] Mark S. Schwartz and Frank Andrasik (eds.), *Biofeedback - a practitioner's guide*, 3rd ed., Guilford, New York, 2003.

[Sal06] Kenneth S. Saladin, *Anatomy & physiology: The unity of form and function*, McGraw-Hill, 2006.

[SAMG97] Michael Singer, Robert C. Allen, Daniel P. McDonald, and John P. Gildea, *Terrain appreciation in virtual environments: Spatial knowledge acquisition*, Technical Report 1014, ADA 325 520, U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA, February 1997.

[SASW96] Victor Spitzer, Michael J. Ackerman, Ann L. Scherzinger, and David Whitlock, *The visible human male: A technical report*, Journal of the American Medical Informatics Association **3** (1996), no. 2, 118–130.

[SB85] Scott N. Steketee and Norman I. Badler, *Parametric keyframe interpolation incorporating kinetic adjustment and phrasing control*, Proceedings of SIGGRAPH '85, 1985, pp. 255–262.

[Sch95] David W. Schloerb, *A quantitative measure of telepresence*, Presence: Teleoperators & Virtual Environments **4** (1995), no. 1, 64–80.

[SECP95] Michael J. Singer, Jennifer Ehrlich, Stephen Cing-Mars, and Jean-Paul Pepin, *Task performance in virtual environments: Stereoscopic versus monoscopic display and head-coupling*, Technical Report 1034, ADA 306 720, U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA, December 1995.

[SFR01] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht, *The experience of presence: Factor analytic insights*, Presence: Teleoperators & Virtual Environments **10** (2001), no. 3, 266–281.

[SGS05] Alistair Sutcliffe, Brian Gault, and Jae-Eun Shin, *Presence, memory and interaction in virtual environments*, International Journal of Human-Computer Studies **62** (2005), no. 3, 307–327.

[SH98] Günter Schreiber and Gerd Hirzinger, *Singularity consistent inverse kinematics by enhancing the jacobian transpose*, Advances in Robot Kinematics: Analysis and Control (Jadran Lenarčič and Manfred L. Husty, eds.), Kluwer Academic Publishers, 1998, pp. 475–483.

[She92] Thomas B. Sheridan, *Musings on telepresence and virtual presence*, Presence: Teleoperators & Virtual Environments **1** (1992), no. 1, 120–126.

[SJW98] Philip N. Sabes, Michael I. Jordan, and Daniel M. Wolpert, *The role of inertial sensitivity in motor planning*, The Journal of Neuroscience **18** (1998), no. 15, 5948–5957.

[SK00] Karan Singh and Evangelos Kokkevis, *Skinning characters using surface oriented free-form deformations*, Proceedings of the Graphics Interface 2000 Conference (Montréal, Québec, Canada), May 15-17, 2000, pp. 35–42.

[Sla04] Mel Slater, *How colorful was your day? Why questionnaires cannot assess presence in virtual environments*, Presence: Teleoperators & Virtual Environments **13** (2004), no. 4, 484–493.

[SLSG01] Hyun Joon Shin, Jehee Lee, Sung Yong Shin, and Michael Gleicher, *Computer puppetry: An importance-based approach*, ACM Transactions on Graphics **20** (2001), no. 2, 67–94.

[SLT86] J. F. Soechting, F. Lacquaniti, and C. A. Terzuolo, *Coordination of arm movements in three-dimensional space. sensorimotor mapping during drawing movement*, Neuroscience **17** (1986), no. 2, 295–311.

[SMH05] Jonathan Starck, Gregor Miller, and Adrian Hilton, *Video-based character animation*, SCA '05: Proceedings of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (Los Angeles, CA), July 29-31, 2005, pp. 49–58.

[SO03] Corina Sas and Gregory M. P. O'Hare, *Presence Equation: An Investigation into Cognitive Factors Underlying Presence*, Presence: Teleoperators & Virtual Environments **12** (2003), no. 5, 523–537.

[SP86] Thomas W. Sederberg and Scott R. Parry, *Free-form deformation of solid geometric models*, Proceedings of SIGGRAPH '86, 1986, pp. 151–160.

[SPB+98] Marius-Calin Silaghi, Ralf Plänkers, Ronan Boulic, Pascal Fua, and Daniel Thalmann, *Local and global skeleton fitting techniques for optical motion capture*, CAPTECH'98: Proceedings of the International Workshop on Modelling and Motion Capture Techniques for Virtual Environments (Geneva, Switzerland) (Nadia Magnenat-Thalmann and Daniel Thalmann, eds.), Lecture Notes in Artificial Intelligence, vol. 1537, Springer-Verlag Berlin / Heidelberg, November 26-28, 1998, pp. 26–40.

[SPCM97] Ferdi Scheepers, Richard E. Parent, Wayne E. Carlson, and Stephen F. May, *Anatomy-based modeling of the human musculature*, Proceedings of SIGGRAPH '97, 1997, pp. 163–172.

[SRC01] Peter-Pike J. Sloan, Charles F. Rose, and Michael F. Cohen, *Shape by example*, SI3D '01: Proceedings of the 2001 symposium on Interactive 3D graphics, 2001, pp. 135–143.

[SRCP02] Francisco J. Seron, Rafael Rodriguez, Eva Cerezo, and Alfredo Pina, *Adding support for high-level skeletal animation*, IEEE Transactions on Visualization and Computer Graphics **8** (2002), no. 4, 360–372.

[SS96] Lorenzo Sciavicco and Bruno Siciliano, *Modeling and control of robot manipulators*, McGraw-Hill, New York, 1996.

[SS98] Kay Stanney and Gavriel Salvendy, *Aftereffects and sense of presence in virtual environments: Formulation of a research and development agenda*, International Journal of Human-Computer Interaction **10** (1998), no. 2, 135–187.

[SS02] Wallace Sadowski and Kay Stanney, *Presence in virtual environments*, Handbook of Virtual Environments: Design, Implementation, and Applications (Kay M. Stanney, ed.), Lawrence Erlbaum Associates, Mahwah, NJ, 2002, pp. 791–806.

[SSUS00] Mel Slater, Amela Sadagic, Martin Usoh, and Ralph Schroeder, *Small-group behavior in a virtual and real environment: A comparative study*, Presence: Teleoperators & Virtual Environments **9** (2000), no. 1, 37–51.

[ST86] J. F. Soechting and C. A. Terzuolo, *An algorithm for the generation of curvilinear wrist motion in an arbitrary plane in three-dimensional space*, Neuroscience **19** (1986), no. 4, 1393–1405.

[ST95] Jianhua Shen and Daniel Thalmann, *Interactive shape design using metaballs and splines*, Proceedings of Implicit Surfaces '95 (Grenoble, France) (M. P. Gascule and B. Wyvill, eds.), Eurographics Association, 1995, pp. 187–196.

[ST04] William T. Stinson and Paul G. Thuriot, *Bulging muscles and sliding skin: Deformation systems for hellboy*, SIGGRAPH 2004: Sketches & Applications, 2004.

[Ste92] Jonathan Steuer, *Defining virtual reality: Dimensions determining telepresence*, Journal of Communication **42** (1992), no. 4, 73–93.

[SU93a] Mel Slater and Martin Usoh, *The influence of a virtual body on presence in immersive virtual environments*, VR '93: Proceedings of the IEEE Virtual Reality Conference (Tony Feldman, ed.), Meckler Limited, London, April 1993, pp. 34–42.

[SU93b] ———, *Representations systems, perceptual position, and presence in immersive virtual environments*, Presence: Teleoperators & Virtual Environments **2** (1993), no. 3, 221–233.

[SUS94] Mel Slater, Martin Usoh, and Anthony Steed, *Depth of presence in virtual environments*, Presence: Teleoperators & Virtual Environments **3** (1994), no. 2, 130–144.

[SUS95] ———, *Taking steps: The influence of a walking technique on presence in virtual reality*, ACM Transactions on Computer-Human Interaction **2** (1995), no. 3, 201–219.

[SvdSKvdM01] Martijn J. Schuemie, Peter van der Straaten, Merel Krijn, and Charles A.P.G. van der Mast, *Research on presence in virtual reality: A survey*, CyberPsychology & Behavior **4** (2001), no. 2, 183–201.

[SW97] Mel Slater and Sylvia Wilbur, *A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments*, Presence: Teleoperators & Virtual Environments **6** (1997), no. 6, 603–616.

[SWK+06] Leonhard Schilbach, Afra M. Wohlschlaeger, Nicole C. Kraemer, Albert Newen, N. Jon Shah, Gereon R. Fink, and Kai Vogeley, *Being with virtual others: Neural correlates of social interaction*, Neuropsychologia **44** (2006), no. 5, 718–730.

[Tay02] Laurie N. Taylor, *Video games: Perspective, point-of-view and immersion*, Master's thesis, University of Florida, 2002.

[TBC98] Paul A. Taylor, Alan Black, and Richard Caley, *The architecture of the festival speech synthesis system*, The Third ESCA Workshop in Speech Synthesis, 1998, pp. 147–151.

[TGB00] Deepak Tolani, Ambarish Goswami, and Norman I. Badler, *Real-time inverse kinematics techniques for anthropomorphic limbs*, Graphical Models **62** (2000), no. 5, 353–388.

[Tha00] Daniel Thalmann, *The virtual human as a multimodal interface*, AVI '00: Proceedings of the Working Conference on Advanced Visual Interfaces (New York, NY), 2000, pp. 14–20.

[Tha06] ———, *personal communication*, June 2006.

[TK03] Barnabás Takács and Bernadette Kiss, *The virtual human interface: A photorealistic digital human*, IEEE Computer Graphics and Applications **23** (2003), no. 5, 38–45.

[TK05] Seyoon Tak and Hyeong-Seok Ko, *A physically-based motion retargeting filter*, ACM Transactions on Graphics **24** (2005), no. 1, 98–117.

[TSC96] Daniel Thalmann, Jianhua Shen, and Eric Chauvineau, *Fast human body deformations for animation and VR applications*, CGI '96: Proceedings of Computer Graphics International, June 1996, pp. 166–174.

[TSIF05] Joseph Teran, Eftychios Sifakis, Geoffrey Irving, and Ronald Fedkiw, *Robust quasistatic finite elements and flesh simulation*, SCA '05: Proceedings of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (Los Angeles, CA), 2005, pp. 181–190.

[TT93] Russell Turner and Daniel Thalmann, *The elastic surface layer model for animated character construction*, CGI '93: Proceedings of Computer Graphics International (Lausanne, Switzerland), Springer-Verlag, Tokyo, 1993, pp. 399–412.

[TT97] John Towell and Elizabeth Towell, *Presence in text-based networked virtual environments or "MUDS"*, Presence: Teleoperators & Virtual Environments **6** (1997), no. 5, 590–595.

[UAW+99] Martin Usoh, Kevin Arthur, Mary C. Whitton, Rui Bastos, Anthony Steed, Mel Slater, and Frederick P. Brooks, *Walking > walking-in-place > flying, in virtual environments*, Proceedings of SIGGRAPH '99, 1999, pp. 359–364.

[UCAS00] Martin Usoh, Ernest Catena, Sima Arman, and Mel Slater, *Presence questionnaires in reality*, Presence: Teleoperators & Virtual Environments **9** (2000), no. 5, 497–503.

[UFN+07] Sebastian Ullrich, Benedikt Fischer, Alexandre Ntouba, Jakob T. Valvoda, Andreas Prescher, Torsten Kuhlen, Thomas M. Deserno, and Rolf Rossaint, *Subject-based regional anaesthesia simulator combining image processing and virtual reality*, BVM '07: Proceedings of the BVM Workshop (Munich, Germany), March 25-27, 2007, pp. 202–206.

[UVW+07] Sebastian Ullrich, Jakob T. Valvoda, Marc Wolter, Gisela Fehrmann, Isa Werth, Ludwig Jäger, and Torsten Kuhlen, *A VR-based virtual hand system for experiments on perception of manual gestures*, Proceedings of the 7th International Workshop on Gesture in Human-Computer Interaction and Simulation (Lisbon, Portugal), May 23-25, 2007.

[VAD+03] Jakob T. Valvoda, Ingo Assenmacher, Christian Dohle, Torsten Kuhlen, and Christian H. Bischof, *NeuroVRAC – a comprehensive approach to virtual reality-based neurological assessment and treatment systems*, MMVR '03: Proceedings of Medicine Meets Virtual Reality 11 (Newport Beach, CA) (J.D. Westwood et al., ed.), IOS Press, Amsterdam, Jan 22-25, 2003, pp. 370–372.

[VAKB04] Jakob T. Valvoda, Ingo Assenmacher, Torsten Kuhlen, and Christian H. Bischof, *Reaction-time measurement and real-time data acquisition for neuroscientific experiments in virtual environments*, MMVR '04: Proceedings of Medicine Meets Virtual Reality 12 (Newport Beach, CA) (J.D. Westwood et al., ed.), IOS Press, Amsterdam, Jan 15-17, 2004, pp. 391–393.

[VdLV+02] Guy Vingerhoets, Floris P. de Lange, Pieter Vandemaele, Karel Deblaere, and Erik Achten, *Motor imagery in mental rotation: An fMRI study*, NeuroImage **17** (2002), no. 3, 1623–1633.

[VGSS04] V. Vinayagamoorthy, M. Garau, Anthony Steed, and Mel Slater, *An eye gaze model for dyadic interaction in an immersive virtual environment: Practice and experience*, Computer Graphics Forum **23** (2004), no. 1, 1–12.

[VKW+04] Jakob T. Valvoda, Torsten Kuhlen, Marc Wolter, Ingo Assenmacher, Martina Graf, Rene Vohn, Ines A. Heber, Martina Piefke, Bruno Fimm, H.-C. Nuerk, Walter Huber, Walter Sturm, and Klaus Willmes, *NeuroMan - Virtuelle Realität für experimentelle Neurowissenschaften [NeuroMan - virtual reality in experimental neurosciences]*, VRAR '04: Proceedings of the 1st Workshop Virtuelle und Erweiterte Realität der GI-Fachgruppe VR/AR (Chemnitz, Germany), Sep 27-28, 2004, pp. 217–228.

[VKW+05] Jakob T. Valvoda, Torsten Kuhlen, Marc Wolter, Claudia Armbrüster, Will Spijkers, Rene Vohn, Walter Sturm, Bruno Fimm, Martina Graf, Ines Ann Heber, Hans-Christoph Nuerk, and Klaus Willmes, *NeuroMan: A comprehensive software system for neuropsychological experiments*, CyberPsychology & Behaviour **8** (2005), no. 4, 366–367.

[VMR+04] Kai Vogeley, M. May, A. Ritzl, P. Falkai, K. Zilles, and Gereon R. Fink, *Neural correlates of first-person perspective as one constituent of human self-consciousness*, Journal of Cognitive Neuroscience **16** (2004), no. 5, 817–827.

[VP01] Spyros Vosinakis and Themis Panayiotopoulos, *SimHuman: A platform for real-time virtual agents with planning capabilities*, Proceedings of the Third International Workshop on Intelligent Virtual Agents (Madrid, Spain) (A. de Antonio, R. Aylett, and D. Ballin, eds.), Lecture Notes In Computer Science, vol. 2190, Springer-Verlag, 2001, pp. 210–223.

[vRvLM05] Arjen van Rhijn, Robert van Liere, and Jurriaan D. Mulder, *An analysis of orientation prediction and filtering methods for VR/AR*, Proceedings of the IEEE Virtual Reality 2005 conference (Bonn, Germany), March 12-16, 2005, pp. 67–74.

[VUK05] Jakob T. Valvoda, Sebastian Ullrich, and Torsten Kuhlen, *Comparative evaluation of user performance for modeling tasks in non-immersive and immersive virtual environments*, 3DUI '05: Proceedings of the IEEE VR 2005 Workshop New Directions in 3D User Interfaces (Bonn, Germany) (Doug Bowman, Bernd Fröhlich, Yoshifumi Kitamura, and Wolfgang Stürzlinger, eds.), March 12, 2005, pp. 95–98.

[VUKB06] Jakob T. Valvoda, Sebastian Ullrich, Torsten Kuhlen, and Christian H. Bischof, *Interactive biomechanical modeling and simulation of realistic human musculature in virtual environments*, BVM '06: Proceedings of the BVM Workshop (Hamburg, Germany), Springer-Verlag, Berlin Heidelberg, March 19-21, 2006, pp. 404–408.

[WAVK07] Marc Wolter, Claudia Armbrüster, Jakob T. Valvoda, and Torsten Kuhlen, *High ecological validity and accurate stimulus control in VR-based psychological experiments.*, EGVE '07: Proceedings of the 13th Eurographics Symposium on Virtual Environments (Weimar, Germany), July 15-18, 2007.

[WB03] Jing Wang and Bobby Bodenheimer, *An evaluation of a cost metric for selecting transitions between motion segments*, SCA '03: Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (San Diego, CA), July 26-27, 2003, pp. 232–238.

[WBL+96] Robert Welch, Theodore T. Blackmon, Andrew Liu, Barbara A. Mellers, and Lawrence W. Stark, *The effects of pictorial realism, delay of visual feedback, and observer interactivity on the subjective sense of presence*, Presence: Teleoperators & Virtual Environments **5** (1996), no. 3, 263–273.

[WC91] Li-Chun Tommy Wang and Chih Cheng Chen, *A combined optimization method for solving the inverse kinematicsproblems of mechanical manipulators*, IEEE Transactions on Robotics and Automation **7** (1991), no. 4, 489–499.

[WDW98] Brenda K. Weiderhold, Renee Davis, and Mark D. Weiderhold, *The effects of immersiveness on physiology*, Virtual environments in clinical psychology and neuroscience (G. Riva and M. D. Weiderhold, eds.), Ios Press, Amsterdam, 1998.

[Wel93] Chris Welman, *Inverse kinematics and geometric constraints for articulated figure manipulation*, Master's thesis, Simon Fraser University, September 1993.

[WFB87] Andrew Witkin, Kurt Fleischer, and Alan Barr, *Energy constraints on parameterized models*, Proceedings of SIGGRAPH '87, 1987, pp. 225–232.

[WG97] Jane Wilhelms and Allen Van Gelder, *Anatomically based modeling*, Proceedings of SIGGRAPH '97, 1997, pp. 173–180.

[WH97] Douglas J. Wiley and James K. Hahn, *Interpolation synthesis of articulated figure motion*, IEEE Computer Graphics and Applications **17** (1997), no. 6, 39–45.

[Whi72] Daniel E. Whitney, *The mathematics of coordinated control of prosthetic arms and manipulators*, Journal of Dynamic Systems, Measurement, and Control **94** (1972), 303–309.

[Win90] Jack M. Winters, *Hill-based muscle models: A systems engineering perspective*, Multiple Muscle Systems (Jack M. Winter and Savio LY. Woo, eds.), Springer-Verlag, 1990, pp. 69–93.

[WJS05] Bob G. Witmer, Christian J. Jerome, and Michael J. Singer, *The factor structure of the presence questionnaire*, Presence: Teleoperators & Virtual Environments **14** (2005), no. 3, 298–312.

[WK88] Andrew Witkin and Michael Kass, *Spacetime constraints*, Proceedings of SIGGRAPH '88, 1988, pp. 159–168.

[WKM96] Yin Wu, Prem Kalra, and Nadia Magnenat-Thalmann, *Simulation of static and dynamic wrinkles of skin*, CA '96: Proceedings of the Computer Animation (Geneva, Switzerland), June 3-4, 1996, pp. 90–97.

[WMCG04] Xiaomao Wu, Lizhuang Ma, Zhihua Chen, and Yan Gao, *A 12-DOF analytic inverse kinematics solver for human motion control*, Journal of Information and Computational Science **1** (2004), no. 1, 137–141.

[WMW86] Geoff Wyvill, Craig McPheeters, and Brian Wyvill, *Data structures for soft objects*, The Visual Computer **2** (1986), no. 4, 227–234.

[Woo00] Ryan Woodland, *Filling the gaps—advanced animation using stitching and skinning*, Game Programming Gems (Mark A. DeLoura, ed.), Charles River Media, 2000, pp. 476–483.

[WP95] Andrew Witkin and Zoran Popović, *Motion warping*, Proceedings of SIGGRAPH '95, 1995, pp. 105–108.

[WP02] Xiaohuan Corina Wang and Cary Phillips, *Multi-weight enveloping: least-squares approximation techniques for skin animation*, SCA '02: Proceedings of the 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, 2002, pp. 129–138.

[WR03] Karl Weilhammer and Susen Rabold, *Durational aspects in turn taking*, Proceedings of the ICPhS 2003 conference (Barcelona, Spain), 2003.

[WS94] Bob G. Witmer and Michael J. Singer, *Measuring presence in virtual environments*, Technical Report 1014, ADA 286 183, U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA, October 1994.

[WS98] ———, *Measuring presence in virtual environments: A presence questionnaire*, Presence: Teleoperators & Virtual Environments **7** (1998), no. 3, 225–240.

[YN03] Katsu Yamane and Yoshihiko Nakamura, *Natural motion animation through constraining and deconstraining at will*, IEEE Transactions on Visualization and Computer Graphics **9** (2003), no. 3, 352–360.

[Zaj89] Felix E. Zajac, *Muscle and tendon: Properties, models, scaling, and application to biomechanics and motor control*, Critical Reviews in Biomedical Engineering **17** (1989), no. 4, 359–411.

[ZB94] Jianmin Zhao and Norman I. Badler, *Inverse kinematics positioning using nonlinear programming for highly articulated figures*, ACM Transactions on Graphics **13** (1994), no. 4, 313–336.

[ZBLN97] Ciyou Zhu, Richard H. Byrd, Peihuang Lu, and Jorge Nocedal, *Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization*, ACM Transactions on Mathematical Software **23** (1997), no. 4, 550–560.

[ZCK98] Qing-hong Zhu, Yan Chen, and Arie Kaufman, *Real-time biomechanically-based muscle volume deformation using fem*, Computer Graphics Forum **17** (1998), no. 3, 20–26.

[ZH02] Victor Brian Zordan and Jessica K. Hodgins, *Motion capture-driven simulations that hit and react*, SCA '02: Proceedings of the 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (San Antonio, TX), July 21-22, 2002, pp. 89–96.

[Zha96] Xinmin Zhao, *Kinematic control of human postures for task simulation*, PhD thesis, CIS, University of Pennsylvania, 3401 Walnut Street, Suite 400A Philadelphia, PA 19104-6228, December 1996.

[ZV03] Victor Brian Zordan and Nicholas C. Van Der Horst, *Mapping optical motion capture data to skeletal motion using a physical model*, SCA '03: Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (San Diego, CA), 2003, pp. 245–250.

Curriculum Vitae

Personal Data

Jakob T. Valvoda was born on June 10, 1974 in Prague

Education

1980 – 1994	Primary and secondary education in Prague, Salem, and Bergheim
June 6, 1994	Abitur – German university entrance qualification in Physics, Mathematics, German, and Philosophy
1994 – 2001	Study of Computer Science with enhanced courses on Medicine RWTH Aachen University
1998 – 1999	ERASMUS scholarship Ecole Polytechnique Fédéral de Lausanne
August 27, 2001	Diploma in Computer Science RWTH Aachen University
2001 – 2007	Doctoral studies in Computer Science RWTH Aachen University

Professional

since August 2001	Research assistant, Virtual Reality group RWTH Aachen University

Publications

A complete list of publications is available at **www.valvoda.com**