Chapter 11

Current Status and Future Development
of Structuring and Modeling Intelligent Appearing Motion
***

Thomas Alexander   Stephen R. Ellis

FGAN - FKIE   *NASA Ames Research Center*
Wachtberg-Werthoven   *Moffett Field, California 94035-1000,*
Germany   *U.S.A.*

"Our nature consists in motion, complete rest is death."
Blaise Pascal (1623-1662), French mathematician,
physicist, and philosopher.

"Eppur si muove—Still, it moves."
Commonly attributed to Galileo Galilei (1564-1642),
Tuscan astronomer, philosopher, and physicist.

The two topics covered by this symposium were intelligent appearing motion and Virtual Environments (VE). Both of these are broad research areas with enough content to fill large conferences. Their intersection has become important due to conceptual and technological advances enabling the introduction of intelligent appearing motion into Virtual Environments. This union brings new integration challenges and opportunities, some of which were examined at this symposium.

This chapter was inspired by the contributions of several of the conference participants, but is not a complete review of all presentations. It will hopefully serve as a basis for formulating a new approach to the understanding of motion within VE.

1. Virtual Environments

Virtual Environments (VE) and Virtual Reality (VR) are now often considered an innovative and natural interface for human-computer-interaction. But what is meant by VE and how does it differ from other human-computer-interfaces?

This section will briefly give some definitions and try to relate VE to other disciplines. The characteristics of VE will be specified, especially with regard to the impact of general and intelligent motion. Finally, a brief preview of future possible development is given.

1.1 What are Virtual Environments?

Virtual Environments (VE) provide new media for communication (Ellis, 1991). They subsume comprehensive technologies for presenting computer-generated scenes to human operators and enabling them to interact with them as if they were real (NATO HFM-021, 2001). VE often make use of multi-modality, including auditory and haptic in addition to visual stimuli. Inclusion of multiple modalities enhances the feeling of subjective presence ("feeling of being there") (Barfield & Furness, 1995; Stanney, 2004). The sense of subjective inclusion or immersion of the user in the computer-generated scene is generally stronger than it is with standard desktop IT-technology. Even when restricted to visual stimuli only, the immersion remains strong due to stereoscopic presentation and simulation of observers' real-time motion through a synthetic environment.

VE systems and displays vary with regard to the proportion of real versus virtual stimuli that they present (Kalawsky, 1993; Milgram et al., 1994). Immersing VE-displays, e.g., head-mounted displays (HMDs), totally exclude real world visual stimuli and present only synthetic stimuli. Other display systems described as Augmented Reality enable presentation of a mixed or augmented environment with virtual and real elements. In this case, synthetic parts may be spatially conformal and appear to be spatially integrated in the real environment (Azuma et al., 2001).

In addition to realistic presentation, interactivity is another basic characteristic of VE. It makes an active exploration and experience possible. This feature also supports the immersion into the virtual scene, especially for understanding of visually complex information. Obviously, interactivity is closely related to dynamics and motion, without these, interaction would not be possible.

However, technical limitations in presenting environmental stimuli and the awareness of being exposed to a VE rather than the real world may affect the interpretation of the environment and cause operators' behavior in a VE to differ from that in the real world. This is especially important in applications like simulator-based training, where training skills and knowledge is the main issue and incorrect training has to be avoided. In general, computer-based training systems are successful for training vehicle drivers, with everything within arm's reach being real and everything out-of-the-window being virtual. However, for training teams with multiple individual viewpoints these systems still need improvement.

FP Brooks' presentation on human motion in *VE for Team Training* illustrated the spectrum of possibilities for team training. For assessing the applicability of VE for training small teams he referred to system effectiveness and influential technological factors of a VE. In his experiments, he focused on observations of human motions and changes of various physiological measures of

effect in a compelling environment. Among others, the technological factors examined were field-of-view, method of travel within a VE, passive haptics, and latency. According to Brooks, a restricted field-of-view caused only limited behavioral differences. Early results assessing differing methods of travel within a VE showed that real walking in a VE closely mimicked reality, whereas an indirect "flying" technique produced motion pathes quite different from motion in the real world. Passive haptics supported the degree of immersion strongly.

1.2 Intelligent Agents and Avatars

Inclusion of virtual, intelligent organisms into the VE can make the system more useful and realistic. This feature is especially important for applications in education and training, when interaction and communication with other, synthetic entities is required. One big advantage of VE in comparison to live training is the more deterministic and totally reproducible setup of training scenarios, and the greater flexibility of the virtual system itself. With today's systems, such a high fidelity of the computer-generated scene can be achieved, that virtual simulation can approach and sometimes even replace live training to some extent.

When VE incorporates motion simulation of anthropomorphic entities or avatars, consistency between the visual appearance of the avatar and its motion becomes essential. A photorealistic rendering with only minimalist, abstract motion simulation is likely to appear incongruous. On the other hand, simple motion simulation for cartoon-like avatars may still seem realistic. Nevertheless, for applications with a high fidelity visual representation, like VE, a high fidelity of the behavior modeling is required.

Realistic motion of anthropomorphic elements within the VE, such as virtual humans, is critical because users of such systems are very sensitive to inaccuracies. From lifelong experience we have gained such detailed knowledge about gestures, facial expressions etc. that we notice even small errors and inconsistencies instantly. Furthermore, we often use motion as an indicator for inferring emotional states, intentions, and goals. Accordingly, slight inaccuracies in motion modeling might therefore easily lead to incorrect inferences about future actions and goals of virtual entities.

There are several approaches for implementing virtual humans with computer-generated behavior into VE. D. Thalmann presented several realistic virtual humans in his presentation on *Intelligent Virtual Humans Behavior.* According to him, the main problem areas were the level of AI within perception and motion control. Yet, only models for generalizing simple types of low-level motion existed, for instance, for walking or reaching. Future applications will require a single, more general model for low-level and high-level motion.

Most virtual human models were developed for computer graphics and related domains. These models look very realistic, especially when visualized as static pictures. But realistic animation has proven difficult and is often implemented off-line by manually programming motion sequences or by controlling the movements by motion recorded from a real actor. In both cases a simplified model is used to visualize the output first, and the final, photorealistic simulation is calculated in a considerably longer period afterwards. Real-time, dynamic photorealistic

rendering is still limited due to available computational resources and performance. Consequently, one still has to trade off photorealism with realistic motion behavior.

There is another discrepancy between modeling behavior on a low (movement) and high (behavioral, cognitive) level. Low-level modeling primarily focuses on single, goal-directed movements, like walking towards or reaching for a target, whereas high-level modeling refers to goals-seeking behavior and goal generation. The models simulating low-level motion behavior are frequently used for workplace design, games, or animation in the movies, and they enable realistic appearing motion for photorealistic human models. A more detailed description of the underlying modeling principles can be found in section 2.2.1 to 2.2.3.

On the other hand, high-level behavior models often lack high quality rendering and visualization. Instead, they model human performance and cognitive processes. Detailed information about them is given in section 2.2.4 and 2.2.5. These models are mainly used for the design of complex working processes including the human-in-the-loop and applications in simulator-based training, especially of vehicle drivers.

In his presentation on *Behavior of Synthetic Entities,* R. Kruk presented approaches for modeling decision-making and more complex behaviors for simulator-based training. It was based on real characteristics of the original system's performance and the decision-making processes actually used in practice. Additional knowledge about, for instance, terrain, culture etc. was included in the model. The behavioral models drove diverse (airborne and grounded) vehicle models within a simulation framework.

Which level or model to choose depends strongly on the specific application. In case, low-level motion is needed, a script may simply control behavior on a higher level to follow special instructional procedures and no high-level control is required. On the other hand, for training organizational processes and simulator-based training of situation management, higher-level modeling or realistic rendering is more important. In this case, low-level movement modeling may be simplified or ignored totally.

Because of the complexity of the whole field, only few multilevel models exist. But recent approaches in defining a common interface between models of different levels are leading towards a more comprehensive approach. They include a high-level behavioral model controlling a low-level movement model. This way the synthetic entity is able to act autonomously in the VE with realistic appearing movements.

The presentation of M. Sierhuis *About Subsumption Architectures in AI* addressed modeling individual behavior and connecting two different behavioral models. The NASA-developed software BRAHMS, a model for describing human behavior, was used to describe and analyze observations of a team in an isolated environment. Subsequently, the model controlled the motion of VE models of the virtual humans. Their behavior was modeled based on the pre-recorded behavioral patterns, and resulting effects of team interaction and actual operations were simulated.

172

Applications like this show that avatars and virtual agents now can be used for system design and work procedure design.

1.3 Adjustable VEs for Intelligent User Support

In addition to animation of the VE content, the capturing and processing of observer's ego-motion is another important aspect of motion in VE. VEs have been characterized before as highly interactive, enabling a natural human-system-interface. Since observer's motions often vary, so do interaction preferences. A simple example of this is handedness. Therefore, it will likely be necessary to customize the VE for each individual. This is especially important because of the close coupling between user actions and system's responses within typical VE-systems.

VE have the capability to enhance reality or to present a modified reality. This is not limited to Augmented Reality or Mixed Reality, but affects each VE application. For instance, navigating through a VE can provide both, global overview (exocentric viewpoint) and local situational awareness (egocentric viewpoint). In his presentation about *Optimizing Viewpoint Control in VE*, P. Milgram described alternative ways for providing spatial situation awareness while navigating and moving through a VE. In contrast to a pure exocentric, global frame of reference, or an egocentric frame of reference, he argued for a spring-damper-coupling between camera and motion combining some of the benefits of both, ego- and exocentric views. In this case the viewpoint was related to both, global overview and personal viewpoint.

But even when limiting the VE to a model of reality, it is not possible to incorporate all aspects of a real environment. Instead, they have to limit interaction capabilities and simulation to application-dependent features and to user-specific needs. For example, it would be senseless to visually simulate events that are beyond users' visual resolution. A future, intelligent VE system could take such limits into consideration and model only the relevant aspects of the environment.

One approach to this would be to include intelligent prediction algorithms for user interactions that take users' and systems' perceptual and motor limitations into account when adjusting parameters such as time-step size in motion dynamics calculations.

At a higher level, an intelligent system would have to include modules to learn the users' natural motions. It would use this knowledge for obviating the need to teach the user new interaction procedures. In this connection, the presentation of K.F. Kraiss on *Hierarchical Structured Nonintrusive Sign Language Recognition* provides an example of the benefits of such an approach since it is self-teaching and adaptive to new signs. By extending the gesture recognition of such systems to continuous gesturing, VE systems would become adjustable to user's natural gesture input. This makes interaction more natural and intuitive.

A further step towards an intelligent system would be to assess the user's preferences, state, and expectations. Out of this, consequences for interactivity and general scene design would have to be inferred. They have to be included in the general scene presentation to minimize additional modeling. The system has to monitor the users' actions continuously and adjust to their actual state. It has to adjust to shift resources dependent on new needs and requirements as a reaction.

The amount of monitoring and system's variability varies widely. Simple systems would include predictors for operator movement behavior to minimize latency effects. An application area, where this feature is crucial, is teleoperation. When using VE for teleoperation, there is high demand on realtime system response, especially when including haptic feedback. For obvious reasons, with an application like teleoperation in surgery this is of even more importance. F. Cardullo addressed the associated problems and proposes possible solutions in his presentation on *Telerobotic Surgery*. It was described how to overcome especially the transport delay problems by using an intelligent system to predict operator actions and reactions on the remote site.

With intelligent modules for the inclusion of further knowledge about system behavior the user performance can be enhanced for more complex teleoperation systems. A Grunwald showed in his presentation on *How Inverse Dynamics Make Users Smart* how supporting the user with adequate clues helps performing complex planning tasks during maneuvering space vehicles under dynamically, counter-intuitive conditions (e.g., the orbit environment with spatial movement restrictions and fuel usage constraints).

Applications like these show the need for including intelligence into the design of future VE systems. Today, VE systems make wide use of e.g., predictive Kalman filters to anticipate users' motion in order to overcome missing positional data and ensure a constant frame rate for rendering (Kalman, 1960). Future complex systems would relate to the computing power and shift computing processes in accordance to their user's preferences. By using operator behavior as input variables, they could serve as prototypes for intelligent VE.

2. Motion and Intelligent Appearing Motion

Motion, the second topic of the symposium, is essential for the perception and understanding of our environment. By simply looking at everyday reality, we can notice that it is never static. The environment itself either causes perceivable motion when objects themselves move, or we cause apparent environmental motion ourselves when we move through the environment.

Sometimes the observed motion appears intelligently driven. This section summarizes several aspects of the appearance and generation of intelligently driven motion. Additionally, different simulations of specifically humanoid motion within computer-graphics are discussed.

Physical motion in Virtual Environments

Physical motion of objects is a topic that has been studied extensively especially in mechanics (Newton, 1687; etc.). It is classically defined as *"…an act, process, or instance of changing place: movement; …an active or functioning state or condition"* (Merriam-Webster Collegiate Dictionary, 2003). Measuring the change of position of an object in a special amount of time quantifies motion leading to the computation of velocity as first derivative of place by time. Motion, however, need not be limited to a change of location, but could also metaphorically describe change of other attributes. Examples of such attributes could be light and color (e.g., due to lighting or shading), personal opinion, or even existence of the content itself in a VE (e.g., objects and sounds could appear or disappear).

Motion is always dependent on factors and physical principles defining constraints, features, and relations.

One example of this is *acceleration,* which describes changes in velocity during time. In a physical environment it is the quotient of force and mass. This simple relationship shows that there are generally two main factors affecting physical motion: Attributes of the static object itself (mass) and additional controlling factors (force). Both together cause a characteristic motion. By knowing all the internal attributes of an object, e.g., mass and other properties, and external factors, e.g., interacting forces, the motion of an object can be described and extrapolated into future. Even more complex motion of intelligent entities can be modeled when there is sufficient knowledge of internal characteristics and external factors available. In this case, internal characteristics are not limited to physical parameters but include psychological aspects as well, e.g., personality and motivation. External factors can either refer to low level of modeling, e.g., forces, obstacles, or to more complex levels, e.g., other persons interacting with the virtual entity. In either case, a complete understanding would make a comprehensive modeling possible.

Internal and external factors are manifold and numerous. As a matter of fact, the variety of possible motions and their combinations goes beyond the scope of modeling. A modeling of a definite motion is difficult or even impossible. Constraints or rules have to be specified in order to minimize the variability and find realistic solutions for motion modeling.

T. Sheridan addressed this issue in his presentation on *Constraint, Intelligence, and Control Hierarchy in Virtual Environments.* He pointed out that utility is the solution of two kinds of equations: on the one hand the objective (utility) function for the goodness of a solution and on the other hand the given constraint equations. This can be applied to language, music, body movements, graphic displays, computer programming and supervisory control, and VE. Constraints that apply in VE are mainly sensory range and resolution of the observer, observation point consistency in space and time, continuity of kinematics in space and time, cause and effect, mechanical impedance interaction with the observer/user, symbolic interchange, and etiquette. He concluded, that, if the many expected constraints are not adhered to, a VE does not appear real or even intelligent.

The reasoning behind the motion and the way it is executed are important for the sensation of realism of the behavior of the content of a VE and, consequently, of the VE itself.

But what makes motion appear intelligent? One approach to answer this question is to refer to the definition of intelligence in biology and to the contribution of behaviorists. W. Kosinski did so in his presentation on the *Pavlovian, Skinner, and Other Behaviorists' Contribution to AI.* Intelligence is defined as the ability for reasoning, imagination, insight and judgment. It requires three fundamental cognitive processes, which are: Abstraction, learning, and dealing with novelty. Behavioral intelligence was only noticed, when the observer saw how that behavior is adaptive. Therefore, intelligence was subjective or "in the eyes of the observer" (Brooks, 1991). Kosinski specifies AI approaches for modeling learning and stresses the importance of the inclusion of AI in complete systems design.

2.2 Simulation of Human Motion

Simulation of human movements has been a research topic in the fine arts, sports medicine and orthopaedics, workplace-design, safety, and, most recently, in computer graphics. A broad variety of computational models already exist for simulating realistic motion and complex behavior. However, most of the approaches used focus only on a single application. The baseline characteristics of such different simulation approaches were addressed in the following presentations at the meeting.

2.2.1 Motion Capture and Animation

In the following, actual approaches and methods of modeling human motion are briefly described. They involve a broad spectrum of diverse application areas like animation in computer graphics, workplace design, and basic research on human behavior.

Motion capture is a direct, straight-forward, and simple way to model motion: By attaching markers to moving objects and measuring the three-dimensional positions during movements, detailed, realistic trajectories are recorded. Reproducing these positions and mapping them on a computer-generated object gives the impression of a very high-fidelity motion simulation.

Nevertheless, this procedure is inadequate for modeling adaptive or intelligent behavior because there is basically no real motion model behind it. But it can serve to generate a database of primitive motions for a more sophisticated model, in which captured data is used to parameterize a motion model for animation of computer-generated objects.

Another way to animate an object is keyframe animation. In this case, captured position data of just a few *keyframes* serve as input. Positions in between are interpolated. With growing complexity of the interpolation algorithms, fewer keyframes are required for a realistic motion.

Such techniques are frequently used in games or movies, where motion follows simple rules or is even script-driven. The primary intention is not a valid model but a realistically appearing output. Most of such models are either special proprietary developments or modules of graphics and animation libraries.

Motions capture and animation are usually limited to movement modeling on a primitive level. They can seldom be applied for simulating complex behavior, especially in combination with goal selection and generation. Instead, they serve as a baseline vocabulary, which is a sort of a database of movements of behavior models of a higher level.

2.2.2 Biomechanic Models

Biomechanic models were originally developed for sports medicine and workplace design. They make use of laws of mechanics and apply them to the human body. The human body is considered as a complex mechanical system (including simple joints of different degrees-of-freedom with static connections between them). Motion is modeled, for instance, by applying

cost functions including forces, torques, load, and (potential and kinetic) energy. This procedure overlaps widely with algorithmic animation, so that a clear distinction between both is not possible.

The original focus of biomechanics was to calculate maximum loads of postures for workplace design and to optimize motion for better performance during energetic work. Biomechanics is often based on real motion capture data, but it generates a more sophisticated model from it. However, due to a large number of factors, the results may look less realistic than the originally captured data.

Like motion capture, biomechanics works primarily for motion modeling on a low level; for modeling behavior, other models have to be applied.

2.2.3 Digital Anthropometric Models

For an ergonomic workplace-design the inclusion of human dimensions has always been essential. At first, physical templates for representing small, medium and large persons were applied for this purpose. With growing importance of computer-aided design media, digital templates were developed and used for the design process. Whereas first digital models were just CAD-versions of the physical templates, not taking into account joint movement limits and applicable only for static applications, today's anthropometric models include complex algorithms for describing human body shape variability and basic human movement behavior. Behavior modeling is mostly based on biomechanic models, and refers only to simple, goal-directed movements like reaching for targets or gait modeling. For modeling of human behavior control and variability, no comprehensive approach exists.

The most common models are JACK (UGS, 2005), RAMSIS and Anthropos (Human Solutions, 2005), and SAFEWORK (Safework, 2005). Both models come with photorealistic rendering and movement modeling capabilities. However, their background and application area varies. The background of JACK is primarily computer graphics and animation. It is frequently used for computer-based training in VE. N. Badler, the model's developer, referred to this model in his presentation on *Meaningful Motion.* The main application area of RAMSIS is the automotive industry. With growing interest in VE in this field, integrating it into VE is currently being evaluated. In this application the designer controls the model by motion capture. For Safeway a similar approach is undertaken. It shows a general trend of inclusion of these detailed models into VEs for various purposes.

2.2.4 Performance Models

The original background of performance models is modeling human reliability and performance for resource and process planning and optimization. They consist of special modules for perception, cognition, and motor reaction. By combining these modules, conclusions about total human-machine-system reliability or performance for highly complex tasks can be inferred. Performance models are widely used to optimize interactive processes of different types.

MIDAS (Man-Machine Integration Design and Analysis Systems) is one of these models. It is an integrated suite of software components developed to aid in the design complex human-machine systems. The goal is to develop an engineering environment, which contains tools and models to assist in the conceptual phase of crewstation development, and to anticipate training requirements (MIDAS, 2005).

These models approach behavior simulation at a higher level. They refer primarily to goal selection and goal generation. With sufficient information and data, it is possible to use them for modeling human decision making during motion planning. In this case, a connection between a high-level performance model and a low-level animation model would make an intelligent model. However, this connection exists only in a very preliminary way in some models, and must still overcome associated technical problems.

2.2.5 Cognitive Models

Cognition is a very complex domain that places great demands on computational modeling. The general problem is that modeling cognition is somehow a modeling of the 2nd order; it is a kind of recursive modeling. This is because a cognitive model has to model a model of the real environment, which was generated by the cognitive system.

Most cognitive models are based on a formal analysis and inferred probabilistic models. They usually serve as a software-tool for solving problems and are based on fundamental knowledge from psychological research and on a variety of different case studies. Cognitive models simulate probabilities of special states and transitions between the states. For instance, working on different tasks sequentially defines a unique order of the tasks. Subjects might choose different orders based on their working strategy. The probabilities of the transitions from one stage to another can be used to model these strategies expressed in Bayesian terms or Hidden Markov Models.

The overall goal of cognitive models is to model human operator behavior with regards to his sensor, cognitive, and motor properties. They can be used for optimization of human-machine-systems at a very early design phase. By modeling operator behavior they can reduce the need for time-consuming experiments and minimize the design iterations for human-machine-system development.

One of these models is the *Atomic Components of Thought—Rational or ACT-R* (Anderson, 1996; CMU, 2005). It serves as a framework for modeling different tasks in a special programming language. This specific model can rely on general assumptions, which are provided by ACT-R or can be specified by the author. ACT-R has been used successfully for producing user models with human-computer interaction that can assess different computer interfaces, cognitive tutoring systems in education, cognitive agents that inhabit training environments and to interpret FMRI data. To some extent an application of such a model for goal-generation and goal-selection might be possible. In this case such a model could be applied to control the goal of a motion of a low-level movement model.

2.3 Perception of Motion and the Impact of Attribution

Despite the fact that motion is often considered to be purely an output process, it is closely related to perception and cognition. Motion is a closed loop, in which perception gives feedback about the effects of motions and movements, resulting again into adjustments for further actions. In a general sense, perception gives us an image and understanding of the environment, and motion is used to respond with appropriate actions. Both areas are closely coupled to each other so that separating them would make it difficult to understand and simulate either of them.

D. Wolpert considered sensory and motor uncertainties forming fundamental constraints on human sensorimotor control. In his presentation on *Uncertainty and Prediction in Sensorimotor Control* he presented the effect of several sensory and motor factors, e.g., noise, on motion. Factors with a major effect on the sensorimotoric system were identified. For instance, Wolpert found 30-70% sensory underestimations of self-produced force, which might be responsible for force escalation in an interpersonal exchange of blows in a kind of a two-person shoving match.

The study of the perception of motion is crucial for generating intelligent appearing motion in VE. For an overall sensation of motion it seems not to be important how exactly low-level motion is modeled, but whether an overall cause or explanation for it can be inferred. In this case the individual disposition, motivation and knowledge of the observer allows him to infer this explanation. He "attributes" to events and these attributions help him to understand even complex situations. The attribution theory, initiated by Fritz Heider in 1958, focuses on the study of the baseline concepts behind this. Causality and intention are two attributions that contribute especially to the perception of motion.

The experiments of Michotte (1946) on causality demonstrate that causal interaction can be perceived with exceedingly simple visual displays. For example, one circle was shown moving from the left to the middle of a display, and a second one moved afterwards from the middle to the right. For different temporal lags or spatial gaps, subjects reported either a continuous motion with one ball hitting the other (causality) or two separate movements (no causality). In his experiments, Michotte found 50 ms to be the threshold value of perceiving causality. With regard to spatial differences, speed was determined to be the all-important factor: With larger speeds the gap could be larger without constraining causality. Notice photorealism was not required to perceive causality.

Metzger (1934) describes another example showing the impact of multimodality on perception of causality. Two identical visual targets moving across each other can be perceived either to bounce off or to stream through each other. By introducing a brief sound at the coincidence most subjects report sensing a bounce. The reason for this sensation and perception is still a topic of actual research. Newtson (1976) supposed that observers seem to unconsciously segment behaviors into actions. Furthermore, expectation leads to inferring actions, preparing and anticipating actions before they take place. If the perceived action is consistent with the expectation, causality is inferred.

In further experiments of Michotte, subjects reported an object either "chasing" and "following" the other, or "guiding" it, dependent on the sizes of the proceeding and succeeding object. These sensations arose due to cognitive attribution during the experience of exocentric motion. The setup of these experiments is illustrated in Figure 1. Notice again the low realism of the setup.
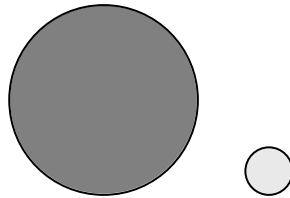

Figure 1. Setup of one of the experiments on causality by Michotte (1946). Subjects reported the small ball guiding the large ball.

In their experiments on the effect of attribution on motion perception, Heider & Simmel (1944) found that even for very abstract motions with geometric shapes human observers tend to build a story around it. In their experiments they used a setup as shown in Figure 2, which consists of an abstract set of two triangles, a circle, and a square that moved around. Subjects watching the scene created a sort of love story around with the circle falling in love with one of the triangles. Haider and Simmel explained these interpretations by the subject's personal attribution.
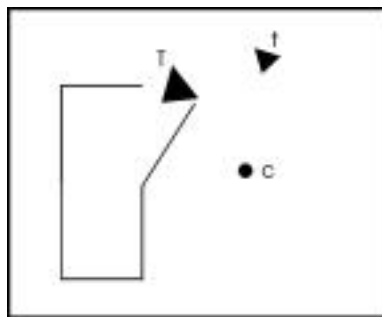

Figure 2. Setup of Heider and Simmel (1944). Subjects reported a love-story of the circle falling in love with one of the triangles and being chased by the other triangle.

Each of these experiments shows that, even without photorealistic appearance and high level, detailed simulation, motion can still appear realistic and intelligent. Such findings contradict an understanding of motion as a purely physical concept. Possible explanations are based on individual perceptual characteristics like attribution. Attribution, as cited before, helps the observer to find an explanation for events and answer the question "why." It results in perceiving basic concepts, for instance, causal links between events.

Causality can be inferred readily if there is a spatial and/or temporal close relationship between two events. *Causal motion* would be an initial level of an additional structure subjectively needed for understanding the appearance of motion. It refers to the relationship of events or movements to each other. Causality is, of course, primarily based on mechanical laws, and can therefore be modeled. However, it is often intuitively inferred, based on experience and knowledge. Because of this attribution, perception of causality is not limited to perception but includes cognition also.

While causality refers to a relationship between single events without more complex interpretation, the theory of attribution hints at higher levels. For a set of events, a goal or an intention is often inferred. This goal is based on the observer's subjective thoughts and feelings and might sometimes be wrong. This level would be the level of *intentional movement*. It refers to the observer's perception and attribution of qualities to the motions.

In his presentation on *The Illusion of Sentience in Virtual Environments,* M. Slater focused on this aspect. In his experiments he found that subjects, especially with anxieties, reacted to virtual humans and interpreted their behaviors as if they were real. This reaction was especially surprising because the simulation of humans was relatively simple and subjects were aware totally of the situation. Especially in setups with anxiety in social situations, the subjects perceived the virtual humans as characters, not as synthetic entities.

These results show that subjective factors like attribution and sensation have a significant impact on the appearance of intelligent motion. Even though there are discrepancies and inaccuracies in geometric or behavioral modeling, a situation might still appear realistic and motion might still appear intelligent because of these perceptual and cognitive factors. Consequently, intelligent motion does not only have an objective dimension, which relates to the scene content and virtual entities, but also a subjective dimension, which relates to the observer. This fact requires additional efforts to specify rules and constraints for modeling intelligent appearing motion.

2.4 Personal Traits and Motion

Causation is not the only subjective attribute that observation of motion may evoke. More complex cognitive attributes may be elicited by the semantic content of a complex scene. For understanding complex scenes different levels of perception and knowledge domains are required. If knowledge from one level is incomplete, knowledge from others may be recruited. An example may be found in the behavior of the entities of Heider's experiments referred to in the previous section, where behavioral explanation spontaneously extended beyond a purely mechanical behavior . In this case social knowledge was recruited to interpret the scene.

Inferring and understanding the semantic content of a situation in a VE can involve hierarchical classifications of objects, their motion, and their behaviors. At the first level, the *primitive categorization,* objects are categorized into animate or inanimate objects based on their behavior. The behavior of animate objects incorporates some sort of intelligence, while inanimate objects show a simple, mechanical behavior only. For animate objects, a second classification may follow. It refers to *primitive psychology,* which includes essential needs of living beings like hunger, thirst, and sleep as the motivation for a special behavior. The third level of categorization refers to *folk-psychology* or naïve psychology. It describes a common understanding of mental

states based on our everyday ascriptions and also includes complex concepts like belief, desire, fear and hope (Fodor, 1987; Goldman, 1993) as motivation. A higher level would introduce *personality traits* to the virtual entity, like extroversion, agreeableness, conscientiousness, emotional stability, and intellect (Wiggins, 1996). Adding these characteristics to behavioral modeling can make the entities' physical actions more transparent to observers and provides the virtual entity with a unique character and behavior.

N. Badler referred to this aspect in his presentation on *Meaningful Motion.* By adding an additional controlling module to a digital human model it was possible to adjust the synthesis of a motion so that specific character attitudes may be observed. In this connection, personality traits and other characteristics are understood as a quality of the motion. The motion itself then still remained goal-directed.

For interaction beyond simple mechanical, *social rules* have to be considered. They include, e.g., occupancy roles, family roles, and stereotypes. A final level for reaching the highest anthropomorphism within a virtual scene would be the inclusion of *high social and political elements* like moral judgment and compassion into the virtual storytelling.

Each of these levels helps an observer understand motion in a virtual scene and behavior of virtual entities. Because of attribution the understanding of the semantic content of the scene is based on the individual knowledge and therefore may differ between observers.

By mixing and applying knowledge from various domains, e.g., psychology and drama, instead of limiting virtual scene generation to pure mechanics, motion and behavior of synthetic entities become much more realistic. This is even true when a single movement or entity appears less realistic, like, e.g., in cartoons or in an abstract setup.

However, inaccuracies or inconsistencies in behavior modeling impair the sensation of naturalistic motion strongly. It has to be pointed out that human observers are very sensitive in detecting deficits due to the extensive knowledge of the relationship between behavior and personality characteristics we have gained during our lifetime.

Relevant knowledge of these basic rules can be found in some surprising fields: Narration and drama. W. L. Johnson's presentation on *Expression, Intention, and Context in Animated Agents* focused on embodied conversational agents and on realistic virtual storytelling. He derived general recommendations for convincing observers that they are observing intelligent agents. The recommendations were originally taken from opera and drama.

While scientific approaches for behavior modeling are often very detailed and complicated, other, more straightforward approaches seem to give good and realistic results. They are often used within the game industry. J. Buchanan referred to possible synergies and differences between these approaches in his presentation on *Video Games: Where Do We Go Now?* The presentation stressed possible applications in edutainment, i.e., education and entertainment, and especially the importance of an effective storytelling for a VE.

3. Comparative Analysis of Descriptions of Intelligent Motion

There are several different ways to classify the large variety of possible motion. A simple, first approach is to differentiate between a pure physical motion on the one hand and intelligently controlled motion on the other. Physical motion would subsume simple movements of dumb, passive objects and low-level movements of animated, active objects. Intelligent motion would focus more on the control behind the movement.

In his opening remarks, S. Ellis proposed such a cybernetic approach for structuring the complex of motion. In this approach, geometry of static objects serves as the basis and includes information about static positions, and restrictions of movable objects or joints. On the next higher level, dynamics are introduced, which refers to changes through time including factors, such as forces and torques. So far each level is based on physical behavior and physical laws. They define the visible output of motion in the VE. With the cybernetic level, control and more complex behavior are introduced. Higher levels relate to goal selection (teleotic level) and goal synthesis (geneotic level).

In this case, behavior is considered a subset of subordinate levels, including goal generation, goal selection, comparison of goal and state, and single movements itself. The approach enables detailed modeling and simulation. A model based on this approach is highly adaptable and would not depend on a specific environment. It would comprise simple movements as well as complex intelligent behaviors.

Current motion models are not able to include this whole spectrum. They are specialized either on modeling lower, i.e., movements, or higher, i.e., behavioral, level motion. Most of them are limited to special applications only and need major transfer modifications for others. But ongoing activities in merging different models show that first approaches for more complex models are coming up.

Nonetheless, the general shortcoming of a purely mechanic/cybernetic approach is that it does not consider the perception of motion and the elicitation of motion attributes. As mentioned before, interrelationships between events (e.g., causality) are often inferred as part of the understanding of more complex behaviors. In these cases personal knowledge is used to fill in the gap due to missing information.

Even simulating simple relationships like causality associated with simulated physical impact gets complicated without considering the perceptual and cognitive processes of the observer. For simulating social relations, it gets much more complicated and complex.

Instead, an alternative description of the relationship between events may be more suitable. This would focus more on the relation between events rather than simulating single, isolated events.

4. An Alternative, Linguistic Approach

Relationships and links between different motions are essential for motion perception and understanding complex intelligent behavior. Physical motion is just one aspect of motion, but causal and intentional movements create the perception and illusion of intelligence within motion. Therefore, the original cybernetic approach for structuring the variability of motion may be extended into another dimension, taking into account explicitly the links and relationships between events and motions that support the concatenation of motion segments.

One possibility would be to include findings in linguistics and semiotics. In this case, motion is considered as a media for communication, instead of being limited to purely goal-directed movements. This extension is important, because VE was previously defined as a new media for communication. A *linguistic approach,* as shown in Figure 3, focuses on a realistic communication between the (virtual) human with the (real) observer/user, on relations between movements, movements and intention, and movements and acting person. In this approach, motion is referred to as a concept of hierarchically constituted entities and patterns, like words and sentences in language. Motion patterns are stored and simply retrieved when needed following syntactic rules.
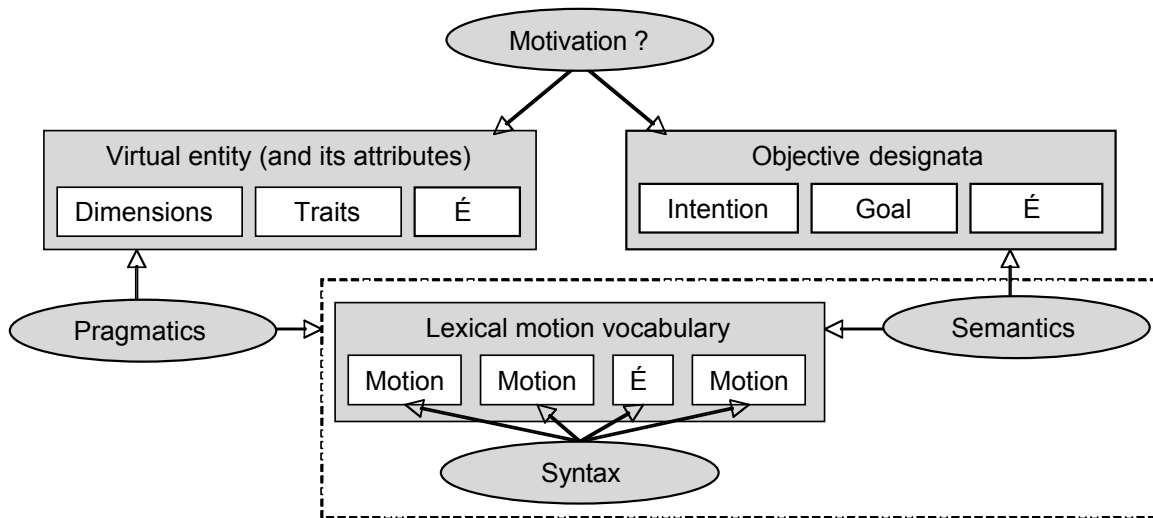
Figure 3. Alternative analytic structure of motion.

At a *lexical* level, it includes basic movements and their relation to static and geometric attributes of the moved entity. Basic kinematics calculated from these attributes are simply retrieved from a movement memory. Interrelations between movements of different parts of the objects, e.g., joint movements, are also described within the lexical level. As a result, the lexical level consists of a small set of simple movements, like words.

The combination of single lexical units to more complex motions happens at the next higher level, which is the *syntactic* level. At this level the relationship between single movements is specified, based on traits of the acting virtual entity. Several single units are combined to form a more complex motion. Transitions between single movements have to be determined to assure continuity. In this level, basic concepts like causality and rules for causality can be specified.

At the *semantic* level, which relates to the designata, movement patterns are correlated to the meaning or overall goal of the motion. The relation between them and the goal of the motion is specified. By this, an observer is able to infer intention and understand even more complex situations.

Intelligence and autonomy are extended on the *conceptual* level. At this level, new goals and concepts for motions are generated. The conceptual level models behavior on a higher level. It has to relate to external factors, like environmental stimuli, as well as internal factors, like motivation and traits of the acting entity.

The relationship between motion and acting person has not been included in this structure. It is handled on the *pragmatic* level, which stands above the other levels. It refers to the relationship between the user and the communications medium.

This approach for structuring motion is based more on the relationship between elements of motion and the overall communicative purpose of the system. It considers motion as a further media to enhance communication between the system and the user. Like storytelling in the movies, drama, or in training concepts, it stresses interactivity and the compositional aspect of complex movement of the VE. By stressing the communication character of motion, a limitation to the communicative-relevant parameters of motion is achieved, whereas simulation of events not relevant to communication can be ignored.

5. Relevance and Future Applications of the Topic

The general trend of motion simulation within VE is for the inclusion of a consistent intelligent behavior for virtual entities, especially virtual humans, on different behavioral levels. Initially, simulation was limited to simple mechanical motion models, e.g., only very simple, robot-like motions and primitive behaviors were modeled. With growing computational resources, simulations include increasingly complex motions today. But still models are very specialized and limited for a single application area only. Only a few of them include more than one level of motion and behavioral modeling. They enable visible realistic movements of the virtual entity as well as selection and generation of underlying goals.

For modeling comprehensive virtual humans in VE only a few approaches exist. This situation is likely to improve in the future because of the development in several application fields.

Today's most common applications of intelligent motion are Virtual Environment systems for education and training. There is a growing demand on realistic training scenarios, which allow training of social skills under different environmental and cultural circumstances. Yet, this is almost exclusively done in physical training areas where the "inhabitants" of a scene are paid

actors and only limited special settings may be trained. This situation is obviously very time- and cost-intensive. Transferring these scenarios into a VE has the big advantage of higher reproducibility and total control of environment variables.

Another application of intelligent motion is related to teleoperation and telepresence. By finding a more efficient way to describe motion pattern with a minimum set of parameters, it would be possible to reduce required transferred information to this set. Ideally, the (intelligent) motion capture system would be able to identify and parameterize motion from a video stream. The consequent parameters would be transmitted to the remote system and control the motion of the remote entity. Especially, when there is only limited bandwidth available, data compression without information loss is required. Such a teleoperation system separates motion information from static (geometric) information, and reduces the amount of data drastically. The transmitted stream would look like a screenplay, referring to objects and their motion.

A more long-term application area is e-commerce and growing use of the World Wide Web. The reason for using anthropomorphic agents or "soft-bots" here is to make boring commerce look more real and to minimize impediments for potential customers. The agent represents an assistant, who aids the visitors of a webpage by helping them find what they are seeking on the page. Speech-understanding software may also be integrated, so that natural language can serve as user input. The agent's reactions include gesture, facial expression and, of course, verbal output. Today's agents are relatively simple and consist of different pictures of synthetic characters, but the development towards virtual, three-dimensional avatars is expected. By applying intelligent motion and "friendly" traits of personality to the virtual human, a friendly and educated assistant may become possible.

A further step would lead to the application of such a virtual human as an intelligent user-interface for a complex system. Complex systems tend to transfer their complexity to the user-interface (UI). For a correct mental model of system status a large amount of user's system knowledge is necessary. By using an anthropomorphic UI, it might be possible to make use of basic inherent social knowledge to enable information transfer in a very realistic and intuitive way.

References

Anderson, J. R. (1996): ACT: A simple theory of complex cognition. American Psychologist, vol. 51, pp. 355-365.

Azuma, R., Baillot, Y., Behringer, R., Feiner, Julier, S. S., and MacIntyre, B. (2001): Recent Advances in Augmented Reality. IEEE Computer Graphics, vol. 21, no. 6, pp. 34-47.

Barfield, W. and Furness, T. A. (1995): Virtual Environments and Advanced Interface Design. New York: Oxford University Press.

CMU (2005): About ACT-R. WWW-document obtained Jan. 05, 2005 from http://act-r.psy.cmu.edu/about/

Ellis, S. R. (1991): Nature and Origins of Virtual Environments: A bibliographical essay. Computing Systems in Engineering, vol. 2, no. 4, pp. 321-341.

Fodor, J. (1987): Psychosemantics: The Problem of Meaning in the Philosophy of Mind. Cambridge, Mass.: MIT Press.

Goldman, A. I. (1993): The Psychology of Folk Psychology. Behavioral and Brain Sciences, vol. 16: pp. 15-28.

Heider, F., and Simmel, M. (1944): An experimental study of apparent behavior. American Journal of Psychology, vol. 13.

Human Solutions (2005): Human Solutions GmbH. WWW-document obtained Jan 5, 2005 from http://www.human-solutions.de/

Kalawsky, R. S. (1993): The science of virtual reality and virtual environments. Wokingham, UK: Addison-Wesley.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. Transaction of the ASME—Journal of Basic Engineering, 82D, pp. 35-45.

Merriam-Webster (2003): Merriam-Webster's Collegiate Dictionary, 11th Edition. New York: Merriam-Webster.

Metzger, W. (1934): Beobachtungen über phänomenale Identität. Psychologische Forschung, vol. 19, pp. 1-60.

Michotte, A. E. (1946/1963): The Perception of Causality. New York: Basic Books.

MIDAS (2005): MIDAS Homepage. WWW-document obtained Jan. 5, 2005 from http://www-midas.arc.nasa.gov/

Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1994): Augmented reality: A class of displays on the reality-virtuality continuum. SPIE, vol. 2351, Telemanipulator and Telepresence Technologies.

NATO HFM-021 (2001): What Is Essential for Virtual Reality Systems to Meet Military Human Performance Goals? Neuilly-sur-Seine: NATO RTA.

Newton, I. (1687): Philosophiae Naturalis Principia Mathematica. Facsimile version by Koyré, A., and Cohen, I. B. (1972), Cambridge, UK: Cambridge University Press.

Newtson, D. (1976): Foundations of attribution: the perception of ongoing behavior. In John H. Harvey, William J. Ickes, and Robert F. Kidd, editors, New Directions in Attribution Research: vol. 1, pp. 223-247, Lawrence Erlbaum Associates.

Safework (2005): Human Modeling Technology. WWW-document obtained Jan. 5, 2005, from http://www.safework.com/

Stanney, K. M. (2004): Handbook of virtual environments. Mahwah, NJ: Lawrence Erlbaum Associates.

UGS (2005): E-factory JACK. WWW-document obtained Jan. 05, 2005 from http://www.ugs.com/products/efactory/jack/

Wiggens, J. S. (1996): The Five-Factor Model of Personality. New York: Guilford Press.