Informed Use of Motion Synthesis Methods

Herwin van Welbergen[1], Zsófia Ruttkay[1], and Balázs Varga[2]

[1] HMI, Dept. of CS, University of Twente, Enschede, The Netherlands
[2] Pázmány Péter Catholic University, Budapest
{welberge,zsofi}@ewi.utwente.nl

Abstract. In virtual human (VH) applications, and in particular, games, motions with different functions are to be synthesized, such as communicative and manipulative hand gestures, locomotion, expression of emotions or identity of the character. In the bodily behavior, the primary motions define the function, while the more subtle secondary motions contribute to the realism and variability. From a technological point of view, there are different methods at our disposal for motion synthesis: motion capture and retargeting, procedural kinematic animation, force-driven dynamical simulation, or the application of Perlin noise. Which method to use for generating primary and secondary motions, and how to gather the information needed to define them? In this paper we elaborate on informed usage, in its two meanings. First we discuss, based on our own ongoing work, how motion capture data can be used to identify joints involved in primary and secondary motions, and to provide basis for the specification of essential parameters for motion synthesis methods used to synthesize primary and secondary motion. Then we explore the possibility of using different methods for primary and secondary motion in parallel in such a way, that one methods informs the other. We introduce our mixed usage of kinematic an dynamic control of different body parts to animate a character in real-time. Finally we discuss motion Turing test as a methodology for evaluation of mixed motion paradigms.

1 Introduction

There are different kinds of human motions, considering the goals and meanings involved, such as communicative and manipulative hand gestures or locomotion. These are all needed in virtual character enhanced applications, like a game. When synthesizing such motions, there are requirements beyond the function: the motion should be realistic, life-like. Motion with the same function repeated by the same person should differ in small motion details.

When observing the entire body, one may notice big motions (like the waving hand), and small motions like balancing with the torso. In this paper we refer to the big motion which is characteristic of the goals and meanings involved as *primary motion*, and the additional details as *secondary motion*. Roughly speaking, when synthesizing the primary motion only, the virtual characters motion can be understood. But without the secondary motions, it will look robot-like and unnatural.

A. Egges, A. Kamphuis, and M. Overmars (Eds.): MIG 2008, LNCS 5277, pp. 132–143, 2008.

From a technological point of view, there are different methods at our disposal: motion capture and retargeting, key-framed or procedural kinematic animation, force-driven dynamical simulation, or the application of Perlin noise. Which method to use, and how to gather the information needed to specify the parameters for the individual methods?

In practice, often the technological considerations – typically, production tools and resources available and the required processing speed – are the factors to justify a single solution. Moreover, as each of the methods are cumbersome, researchers basically concentrate on one of the approaches and improving its algorithms. This is characteristic of the rather disjoint work going on in the 'mocap world' and the 'motion generation' world. The dichotomy is also to be noticed when looking at the application domains. In medical applications one can find high quality dynamical simulation to study walking under different conditions, in movies the quality of (virtual human) motion is high due to using mocap, while in the conversational agents domain, procedural models are used.

The main topic of our paper is to investigate how different techniques can be used to synthesize both primary and secondary motions. Using combinations of these techniques, we aim to come close to the illusion of full realism of the behavior, while also being responsive to the environment the motion is executed in and reactive to the gamer's behavior in real time. We consider informed usage of motion synthesis methods, in two interpretations of 'informedness'. First, one can use motion capture data to inform the (models for) generative methods, particularly, to decide about the primary and secondary motions. Second, one can use multiple methods together for different body segments for primary and secondary motions in such a way that one method informs other(s), see Fig. 1.

We apply our motion synthesis algorithms on our virtual conductor [1] and a reactive virtual trainer [2]. The virtual conductor can conduct human musicians in a live performance interactively. Using knowledge of the musical piece (tempo, volume, the different voices, etc...) he leads the musicians through the piece and corrects them when certain types of mistakes occur. The Reactive Virtual Trainer is capable of presenting physical exercises that are to be performed by a human, monitoring the user and providing feedback at different levels. Both applications require real-time adaptation and precise control of the timing of motion. The primary motion of the conductor consists of conducting gestures executed with one or both arms and/or the head. The secondary motion of the conductor is modeled as balancing motion on the lower body. For the trainer, the primary and secondary motions are exercise depended.

In the next section, we introduce each motion synthesis method shortly, and give an overview of state of the art of using multiple methods in one framework. Then, in section 3, we discuss how motion capture can be used to identify joints to be modeled by different methods, and how to tune the parameters of these methods. In section 4, we discuss the potentials of mixing multiple methods to animate different parts of the body. We demonstrate the idea by an example from our ongoing work, showing mixed usage of kinematic an dynamic control of different body parts to animate a character in real-time. The paper finishes

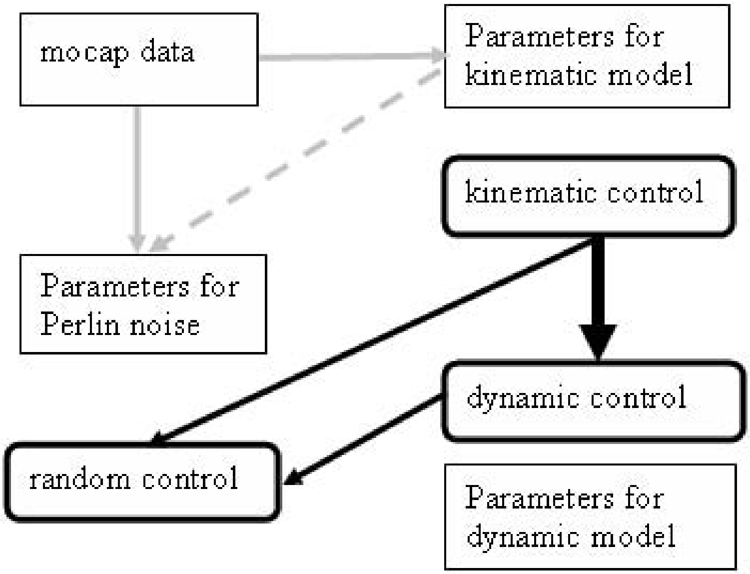

Fig. 1. The three procedural control paradigms, gray arrows showing off-line parameter acquisition based on information from other methods. The other arrows show the possibility for information exchange for real-time parallel tuned usage of multiple methods. The thick arrow indicates the novel method discussed we implemented.

by discussing the motion Turing test, as the evaluation paradigm we plan to use in the future to evaluate mixed motion models.

2 Related Work

Current computer animation techniques make use of kinematic and/or dynamic techniques for motion generation.

Motion editing uses recorded motion as a basis and creates variations on this motion to adapt it to specific needs. This kinematic technique employs the detail of captured motion but only allows very small adaptations. It is relatively simple to use slight motion adaptations to adhere to strict time constraints [3]. Typically, the actor that is motion captured does not have the same limb size as the VH that is to be animated. Therefore, it is necessary to retarget the motion of the actor to make it work on this VH [4]. Sensor noise, motion retargetting and motion adaptation causes physical anomalies on edited motion. One of such anomalies is footskate: the VH's foot slides on the floor after the VH plants it, rather than remaining tightly in place [5]. Another typical artifact of motion editing is the violation of balance, which results in a VH that not seems to be situated in the virtual world but rather seems glued in front of it [6].

Procedural animation is a kinematic technique that uses mathematical formulas that prescribe the change of posture over time, given a set of movement parameters values. Procedural animation can be used to directly control the rotation of joints [7]. A typical application is at a slightly higher level, for generating gestural behavior: the movement path of hands through space is defined mathematically, and the appropriate arm joints are computed automatically [8, 9, 10]. This approach is very adaptable: it offers precise timing and real-time parameterization using a large number of parameters. However, it is hard to incorporate movement details such as those found in motion recordings into the mathematical formulas that steer procedural motion. Physical believability has to be authored explicitly in the formulas. Typically, only a subspace of parameter values yields physically realistic motion.

In *dynamic simulation*, the VH is controlled by applying torques on the joints and a force on the root. A physical simulator then moves the VH's body using Newtonian dynamics, taking friction, gravity and collisions into account. The process of finding accelerations of rigid bodies, based on forces and torques is called *forward dynamics*. A dynamic controller can provide control in real time [11], if the dynamical model of the human is kept simple. Such a controller calculates the torques that will move the VH toward a desired state, based on a heuristic movement model. The input to such a controller is the desired value of the VH's state, for example desired joint rotations or the desired position of the VH's center of mass (CoM). The goal of the system is to minimize the discrepancy between the actual and desired state. The controller can deal with perturbations. For example, if balance is disturbed by a push, the controller automaticly guides the VH back to a balanced pose. However, in general, the controller can not predict when the desired state is reached, or if it will be reached at all. Motion generated by simulation lacks the kind of detail that is seen on captured motion. While the motion is physically correct, this alone is often not enough for movements to be human-like. Therefore, dynamic simulation is mainly used to generate human motion that is physically constrained, and in which interaction with the environment is important, such as motion by athletes [11], stunt men [12], or people falling over [13].

2.1 Combining Dynamic and Kinematic Animation

The previously introduced animation paradigms each have their own advantages, so it would be beneficial to combine them in a single motion system. Existing hybrid systems [12, 13] typically switch between motion editing and dynamic simulation, depending on the current situation's needs.

However, in many situations, the different positive features of the generation paradigms are needed at the same time, but at different body parts. For example, while standing and gesturing, our lower body is in contact with the ground and has to support the upper body, showing a physically realistic balanced pose which would be nice to model with dynamic simulation, while arm gestures require motion detail and tight time synchronization to speech, which would be best achieved by kinematic motion.

In [14] it is shown how inverse and forward dynamics can be combined in a single dynamic animation system, assuming that either the joint acceleration or the joint torque is known for each joint, at each time frame. A similar approach is commonly used in biomechanics to visualize the biomechanical movement model of interest on some body parts (using joint torques), enhanced with known motion on other body parts (using kinematic motion) [15]. Our work on mixed kinematics and dynamics revives the ideas in [14] and shows how they can be applied in an efficient real-time motion system.

2.2 Tracking Kinematic Motion

Motion tracking [16], uses a controller to compute the torque on each joint. The desired state for this controller is the desired rotation of the joint, as specified in by the kinematic data (e.g. mocap). Motion capture noise, retargetting errors, tracking errors and environmental changes can easily disturb the balance of a character whose full body is animated using a tracking controller. Therefore an explicit dynamic balancing model is needed. Because tracking makes use of dynamic controllers, the motion generated by these methods has a time-lag with that specified in the motion capture data. Tracking controllers are computationally more demanding than our mixed kinematic/dynamic method, but provide collision with the environment for the tracked body parts.

3 Modeling Informed by Motion Capture

In the recent years, we have been using motion capture technology to analyze different aspects of hand gestures and full body physical exercise motions [17,18]. We use the results to inform the different modeling paradigms used to synthesize motions. The domain of our investigations are physical exercises for a Reactive Virtual Trainer application [2] where physical realism is of major importance, repetitive motions like clapping [17] or conducting [1], and some communicative gestures like hand-shaking [19]. In all cases, we recorded several samples of the same gesture, by the same subject.

3.1 Identification of Primary and Secondary Joints

The primary and secondary joint motions were identified by looking at the standard deviation of the coordinates of marker positions for each joint, normalized by maximum joint extension. For each motion, the joints are classified as primary, secondary or unused in an automatic way. Currently, the two separating thresholds are to be given manually as a first step, but fully automatic clustering may be used in the future. For all but one recording of rhythmic physical exercise samples, this method identifies the 1-6 primary joints faithfully. That is, the judgment of the classification system corresponds to the judgment of an expert.

3.2 Kinematic Model Based on Motion Capture

In ongoing work, we wish to use the captured time functions of the motion of primary joints to detect certain kinematic characteristics of the motion, and to define parameters for kinematic models. As of motion phases of a hand gesture or exercise, we adopt the terminology used for communicative gestures [20], identifying *preparation*, *pre-stroke hold*, *stroke*, *post-stroke hold* and *retraction*. As an exercise is to be performed in a repetitive way, the retraction is identical with the preparation of the next exercise. By analyzing the captured data, we are to identify:

1. tempo
2. timing of the preparation, hold and stroke phases
3. amplitude, expressed by the maxim values of the displacement of certain body part
4. synchrony

By having recorded tempo variants of the same motion, we also get information of extreme and default tempi, and the correlation between tempo and amplitude, and tempo and timing of stages. This quantitative information is to serve as a basis for kinematic models. The timing information of the phases is gained by identifying points in the motion on which the speed crosses a predefined minimum threshold (see Fig 2).

Synchrony is to identify if (a subsets of) the joints involved move in sync. This is to be decided by analyzing the stroke times of the different joints. In case of physical exercises, a few synchrony patterns are common [2].

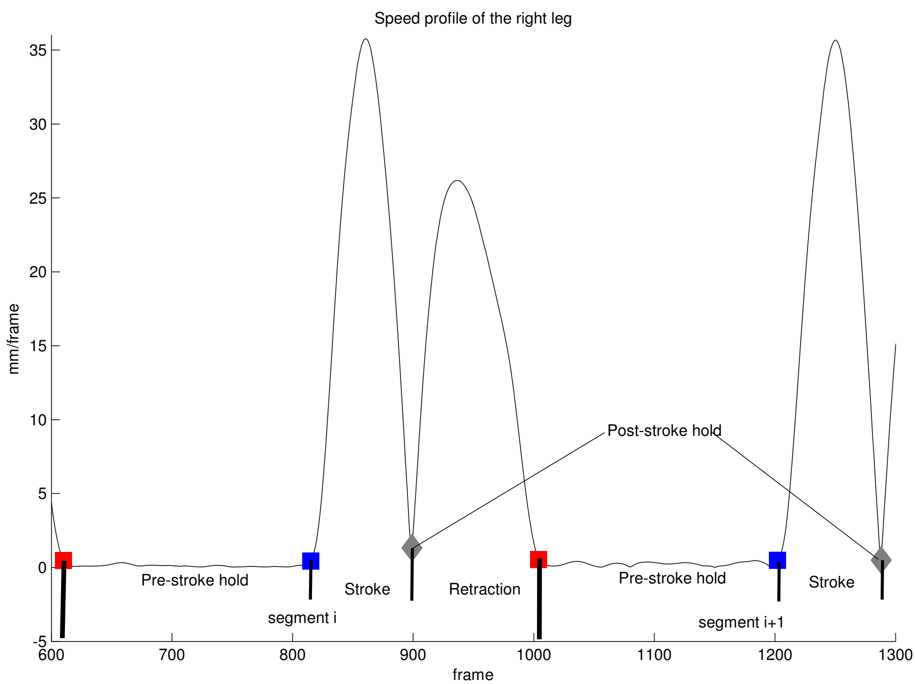

Fig. 2. Motion phases are detected from the speed profile

The physical and procedural models themselves are typically created on the basis of models from biomechanics or behavior science, rather than directly basing them off motion capture. Motion capture information is used to identify (among others) timing, amplitude and synchrony parameters and to select a parameterized model that fits the observed motion. The parameters that steer such a model are designed to be intuitive for motion authors, but are often related. Motion capture can serve as a way to find dependencies between these parameters. For example, we have shown that the movement path of the hand decrease linearly with the tempo in a clapping task [17]. A change of one parameter value then changes all parameters values that are related to it. If a motion generation process specifies the value of more than one parameter, conflicts might arise. These conflicts can be solved in several ways, for example by finding some kind of 'best fit' of parameters values, weighted by their importance.

3.3 Analyzing Variability and Symmetry

We gathered qualitative and quantitative information on the variation between repeated motions with the same function. Even in case of physical exercises performed by expert, there is a small variation in performance. Similar considerations hold for symmetry (see Fig. 3). The amount of variation can depend movement parameters. For example, Fitts' law [21] states that quick pointing movements are neceserely less precise than slower pointing movements.

Small, but consistent phase differences can occur in symmetrical movement. For example, we have shown that right-handed subjects 'lead' a clapping movement with their right hand [17]. The mean phase difference and variation of phase difference can vary with movement parameters. For example, we found that the standard deviation of the phase difference between hands in a clapping task increases with tempo [17].

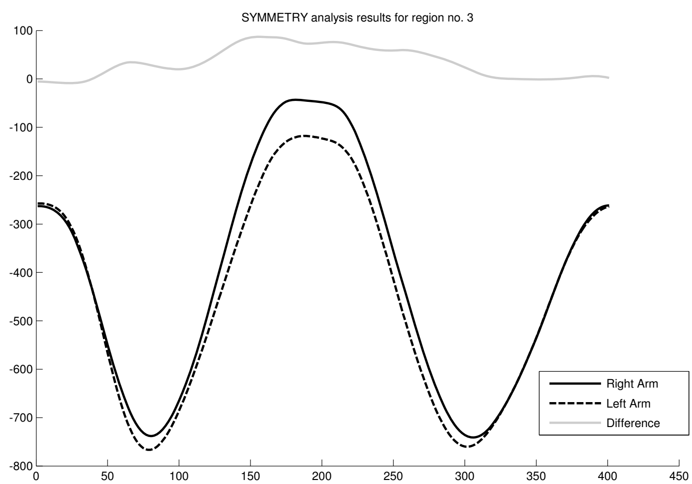

Fig. 3. Symmetry analysis of an exercise motion

4 Using Dynamic Simulation and Kinematics in Parallel

In [22], we show how to simultaneously combine dynamic motion with kinematicaly driven motion (like motion capture, procedural motion). In case of the conductor, the body is segmented in one physically steered group of joints P, located at the lower body and three chains of kinematicly controlled joints, $K_{rightarm}$, $K_{leftarm}$ and K_{head} rooted at the right shoulder, left shoulder and neck respectively (see Fig. 4). The movement of the lower body is steered by the dynamic balance controller, described in [11].

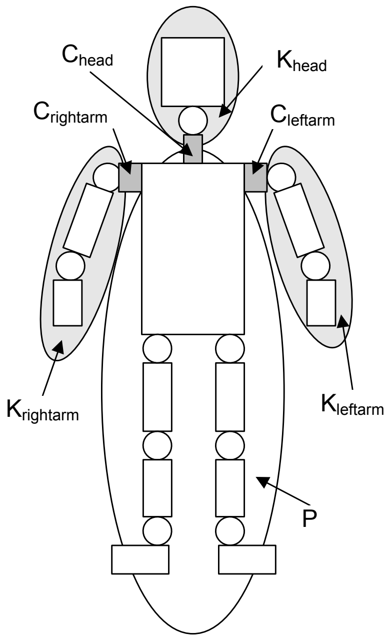

Fig. 4. Kinematically and dynamically steered joints in the conductor

The torque the kinematic chains exert at connectors ($C_{rightarm}$, $C_{leftarm}$ and C_{head}) is calculated using inverse dynamics. The reactive torques are then applied to the lower body. To calculate this torques, we need to know the velocity and acceleration of the connectors. However, the velocity and acceleration of a connector is dependent on the movement of all joints in the body, and can only accurately be calculated by an algorithm that takes the accelerations of all joints in K and the torques of all joints in P into account simultaneously. Rather than directly calculating the acceleration of the connectors at the current frame, we approximate torques that each K exerts on P using the acceleration and velocity of connector at the *previous* frame. This results in a slightly more efficient algorithm than first proposed in [14], both in terms of calculation time and memory bandwidth. More importantly, it allows us to make use of any existing real-time forward dynamics engine for mixed motion generation.

Fig. 5 shows the combination of our physical balance model with a procedurally generated large arm swing. The results of our system are subtle and therefore hard to capture on a series of images. We refer the interested viewer to the demonstration videos [1] to see our system in action with a combination of motion captured arm movements or procedural conducting gestures and the balance model.

5 Further Work

5.1 Generating Variability

In some of our applications (conductor, fitness trainer) the same movement is to be repeated many times. If such movement looks exactly the same for repetition, the believability of our VH is destroyed. We intend to use one of the existing approaches from the literature [7,23,24] to model motion variability. Our motion capture analysis provides us with information on the amount of variability in each motion, on what joint or motion parameter this variability is to be applied and how parameter values affect its size.

5.2 Timing of Motion Phases

We have timing information on each movement phase in our exercise motion. We intend to explore the relation of the duration of each phase length with other parameters, such as movement tempo or amplitude. In our preliminary work on the analysis of clapping [17], we found an invariance in the relative timing of the different phases of clapping movement: no matter the tempo, the same percentage of time was spend in each movement phase. These percentages were slightly different for different subjects.

5.3 Evaluating Movement Models

VHs usually do not have a photo-realistic embodiment. Therefore, if the naturalness of VH animation is evaluated by directly comparing moving humans with a moving VH, the embodiment could bias the judgment. To remedy this, motion captured human movement can be casted onto the VH, thus showing the different motions on identical virtual body. This motion is then compared with generated animation. Typically this is done in an informal way. A *motion Turing Test* [25] could be used to do this more formally. In such a test, subjects are shown generated movement and similar motion captured movement, displayed on the same VH. Then they are asked to judge whether this was a 'machine' moving or a real human.

However, such a direct human judgment is not sufficient to measure the naturalness of motion. Even if a certain movement is judged as natural, an unnatural artifact that is not noticed consciously can still have a social impact [26]. Characters with an unnatural motion can be evaluated as less interesting, less pleasant,

[1] Available from: http://www.herwinvanwelbergen.nl/phd/mixed/mixed.html

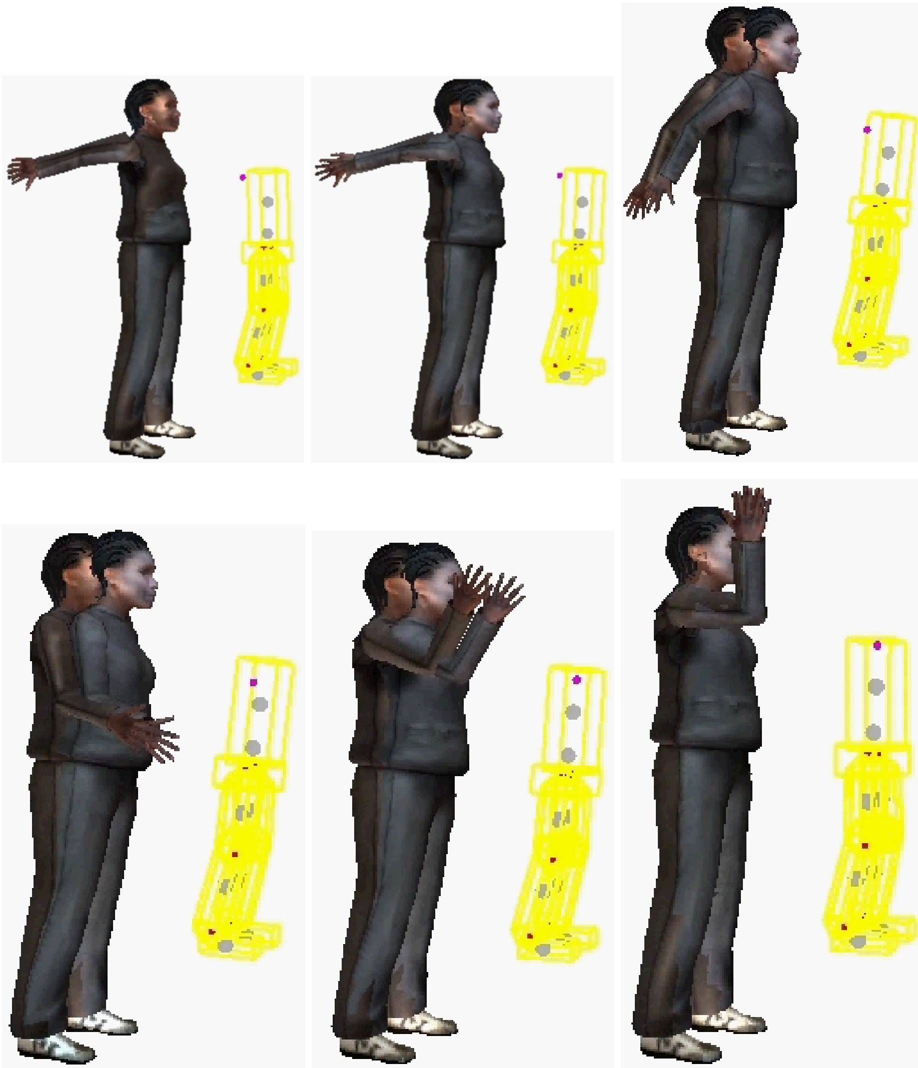

Fig. 5. Mixing a kinematic arm swing with dynamic balancing. The right side of each picture shows the visualization of the physical model of the lower body of the VH. The left side shows a VH with physical lower body control, overlayed with one that does not move the lower body. The counter-clockwise angular acceleration of the arm in the upper three frames causes a clockwise torque on the trunk, which makes the upper body bend forward. The clockwise angular acceleration of the arm on the lower three frames causes the body to bend backward.

less influential, more agitated and less successful in their delivery. So, while a VH Turing test is a good first indication of naturalness (at least it looked human-like), further evaluation should determine if certain intended aspects of the motion are delivered. Such aspects could include showing emotion, enhancement of the clearness of a spoken message using gesture, showing personality, etc.

Our simultaneous motion mixing techniques allow us to let a motion model steer a part of the body, and let a motion capture recording steer the remaining parts. We can then use the motion Turing test to compare the motion generated by the movement model combined with motion capture with the same motion generated solely by motion capture. In a similar way, we can test if a certain aspect of motion is important for naturalness, by using a model that either removes this aspect, or replaces it by noise. In a later stage, we plan to use this approach to test combinations of motion models that were evaluated to work well in isolation.

Further on, we would like to explore the parameter space that yields natural looking procedural or dynamical motion. If the parameter set is small, this natural parameter space can be identified by having subjects directly set and evaluate the possible parameter values [23].

Acknowledgments

This research has been supported by the GATE project, funded by the Netherlands Organization for Scientific Research (NWO) and the Netherlands ICT Research and Innovation Authority (ICT Regie). The motion capture recording and analysis was carried out in the framework of a Short Mission supported by the EU Action COST2102.

References

1. Reidsma, D., Nijholt, A., Bos, P.: Temporal interaction between an artificial orchestra conductor and human musicians. ACM Computers in Entertainment (to appear, 2008)
2. Ruttkay, Z., van Welbergen, H.: Elbows higher! performing, observing and correcting exercises by a virtual trainer. In: Proceedings of Intelligent Virtual Agents, Tokyo, Japan (to appear) (September 2008)
3. Witkin, A., Popovic, Z.: Motion warping. In: SIGGRAPH, pp. 105–108. ACM Press, New York (1995)
4. Gleicher, M.: Retargetting motion to new characters. In: SIGGRAPH, pp. 33–42. ACM Press, New York (1998)
5. Ikemoto, L., Arikan, O., Forsyth, D.A.: Knowing when to put your foot down. In: Proceedings of Interactive 3D graphics and games, pp. 49–53. ACM Press, New York (2006)
6. Tak, S., Song, O.-Y., Ko, H.-S.: Motion balance filtering. Computer Graphics Forum 19(3) (2000)
7. Perlin, K.: Real time responsive animation with personality. IEEE Transactions on Visualization and Computer Graphics 1(1), 5–15 (1995)
8. Chi, D.M., Costa, M., Zhao, L., Badler, N.I.: The EMOTE model for effort and shape. In: SIGGRAPH, pp. 173–182. ACM Press/Addison-Wesley Publishing Co., New York (2000)

9. Neff, M., Kipp, M., Albrecht, I., Seidel, H.P.: Gesture modeling and animation based on a probalistic recreation of speaker style. Transactions on Graphics (to appear, 2008)
10. Hartmann, B., Mancini, M., Pelachaud, C.: Implementing expressive gesture synthesis for embodied conversational agents. In: Gibet, S., Courty, N., Kamp, J.-F. (eds.) GW 2005. LNCS, vol. 3881, pp. 188–199. Springer, Heidelberg (2006)
11. Wooten, W.L., Hodgins, J.K.: Simulating leaping, tumbling, landing, and balancing humans. In: ICRA, pp. 656–662. IEEE, Los Alamitos (2000)
12. Faloutsos, P., van de Panne, M., Terzopoulos, D.: The virtual stuntman: dynamic characters with a repertoire of autonomous motor skills. Computers & Graphics 25, 933–953 (2001)
13. Zordan, V.B., Macchietto, A., Medina, J., Soriano, M., Wu, C.C.: Interactive dynamic response for games. In: Proceedings of the SIGGRAPH symposium on Video games, pp. 9–14. ACM Press, New York (2007)
14. Isaacs, P.M., Cohen, M.F.: Controlling dynamic simulation with kinematic constraints. In: SIGGRAPH, pp. 215–224. ACM Press, New York (1987)
15. Otten, E.: Inverse and forward dynamics: models of multi-body systems. Philosophical Transactions of the Royal Society 358(1437), 1493–1500 (2003)
16. Zordan, V.B., Hodgins, J.K.: Motion capture-driven simulations that hit and react. In: Proceedings of the Symposium on Computer Animation, pp. 89–96. ACM Press, New York (2002)
17. van Welbergen, H., Ruttkay, Z.: On the parameterization of clapping. In: Gesture Workshop, Lisbon, Portugal (July 2007)
18. Varga, B.: Movement modeling and control of the reactive virtual trainer. Master's thesis, Peter Pazmany Catholic University, Dept. of Information and Technology, Budapest, Hungary (2008)
19. Ruttkay, Z., van Welbergen, H.: Let's shake hands! on the coordination of gestures of humanoids. In: Artificial and Ambient Intelligence, Newcastle University, Newcastle upon Tyne, UK, pp. 164–168 (April 2007)
20. McNeill, D.: Hand and Mind: What Gestures Reveal about Thought. University of Chicago Press, Chicago (1995)
21. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology 47(6), 381–391 (1954)
22. van Welbergen, H., Zwiers, J., Ruttkay, Z.: Real-time animation using a mix of dynamics and kinematics. In: Symposium On Computer Animation (submitted, 2008)
23. Bodenheimer, B., Shleyfman, A.V., Hodgins, J.K.: The effects of noise on the perception of animated human running. In: Computer Animation and Simulation (September 1999)
24. Egges, A., Molet, T., Magnenat-Thalmann, N.: Personalised real-time idle motion synthesis. In: Pacific Graphics, pp. 121–130. IEEE Computer Society, Washington (2004)
25. Hodgins, J.K., Wooten, W.L., Brogan, D.C., O'Brien, J.F.: Animating human athletics. In: SIGGRAPH, pp. 71–78. ACM Press, New York (1995)
26. Reeves, B., Nass, C.: The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press, New York (1996)