Virtual Rap Dancer: Invitation to Dance

D. Reidsma
Human Media Interaction
University of Twente
PO Box 217, 7500 AE
Enschede, the Netherlands
d.reidsma@ewi.utwente.nl

R.Rienks
Human Media Interaction
University of Twente
PO Box 217, 7500 AE
Enschede, the Netherlands
r.j.rienks@ewi.utwente.nl

A. Nijholt
Human Media Interaction
University of Twente
PO Box 217, 7500 AE
Enschede, the Netherlands
a.nijholt@ewi.utwente.nl

H. Hondorp
Human Media Interaction
University of Twente
PO Box 217, 7500 AE
Enschede, the Netherlands
g.h.w.hondorp@ewi.utwente.nl

R. Poppe
Human Media Interaction
University of Twente
PO Box 217, 7500 AE
Enschede, the Netherlands
r.w.poppe@ewi.utwente.nl

CHI 2006, April 22–27, 2006, Montréal, Québec, Canada.
ACM 1-59593-298-4/06/0004.

Abstract
This paper presents a virtual rap dancer that is able to dance to the beat of music coming in from music recordings, beats obtained from music, voice or other input through a microphone, motion beats detected in the video stream of a human dancer, or motions detected from a dance mat. The rap dancer's moves are generated from a lexicon that was derived manually from the analysis of the video clips of rap songs performed by various rappers. The system allows for adaptation of the moves in the lexicon on the basis of style parameters. The rap dancer invites a user to dance along with the music.

Keywords
Entertainment computing, multimodal interaction, embodied agents

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
We look at a hardly studied phenomenon of rap performances: the series of gestures and bodily movements that are made by rappers while performing. It is not meant as a study of a rapper's movements and gestures from the point of views of the lyrics. Our study consists of observing various rappers with the aim of distinguishing characteristic movements and regenerating them in an interactive and entertaining embodied agent, a virtual rapper. This virtual rapper has been designed in such a way that it can interact

with a user. In a first version of the rapper environment there is just music input to the rapper that allows him, based on the audio beats, to perform sequences of rapper movements. The movements have been obtained from studying video clips of performing rap artists. In the second version of our system we introduced interactivity. Rather than having the rapper perform on stored audio, we allowed the rapper to perform on input obtained from a microphone and input obtained from a camera. In both cases "beats" (audio and visual) are extracted from the input and these beats guide the performance of the rapper. In the third version of the rapper system, presented here, we have added a dance mat to keep track of the feet of the human dancer and we have added interaction behavior to the rapper that makes it possible to have the rapper display and recognize cues that help to invite a passerby or a user to get involved with the rappers interests and activities. In particular, we let the virtual rapper display behavior that invites human partners to join him in a dancing activity.

Music, dance and interaction: Some previous research

There have been many research and art projects where movements of dancers or players are captured by motion capture sensors or cameras. The movements, as well as other input signals (e.g., speech, facial expressions) can be analyzed and mapped on avatars and other (semi-)autonomous agents. This mapping or re-generation can be done in real-time, allowing applications as interactive theatre and virtual storytelling (see [1,2] for some pioneering work), or off-line, allowing more advanced graphics and animation, but less direct interaction. These latter applications are for example the simulation of baroque dances [3] with virtual characters and the generation of new dances composed from extracted primitive dance movements and newly learned choreographies.

There are also many examples of music interfaces. Previously we have looked at recognizing and distinguishing percussion instruments and music visualization by an embodied performer [4]. Goto et al. [5] introduced an embodied agent that enables a drummer and a guitarist connected through Ethernet not only to musically interact with each other, but also implicitly through the animations of a single character that performs dance motions chosen on the improvisations of the guitarist, timed to the drum. A jam session system that allows a human guitarist to interplay with virtual guitar players can be found in [6]. Different reaction models for human players can be obtained and imitated. An example of a system that extracts acoustical cues from an expressive music performance in order to map them on emotional states is CUEX (CUe EXtraction) [7]. This system has been interfaced with Greta, an embodied conversational agent in which the emotional states obtained from the music are transformed to smoothly changing facial expressions providing a performer with visual feedback [8]. In [9] and [10] a system is described in which musical features are extracted (pitch, specific chords) from a digital piano and microphone, and responsive behavior in synthetic characters is displayed in real-time, influenced by a cognition layer relating music and emotion. Well known research on extracting emotions from dance and body movements is performed by Camurri (see e.g., [11]).

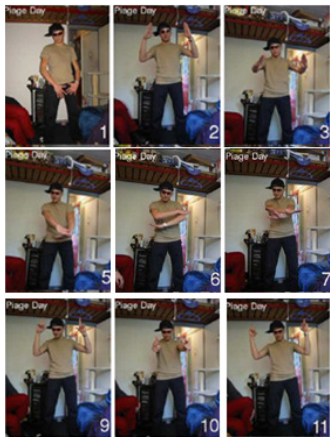

Figure 1: The Kris Kros move

Detailed verbal descriptions for these movement sequences have been made in order to allow corresponding animations of an avatar. For example, the Bitching ho no! sequence is a movement sequence where only the right arm is used. First the right hand is moving at head height from the right of the head to the head, making three duck quack movements, next the right hand index finger turns around on the right hand side of the head.

Analysis of rap gesture sequences

Various rap-video clips have been analyzed, and characteristic movements have been extracted. Fourteen rap movement sequences have been distinguished in the clips and have been selected for a database from which our virtual rapper is fed. All movement sequences were studied, exercised and photographed, and translated to a key frame animations database. In Figure 1 we illustrate the movement sequence ('Kris Kros'). All sequences, with detailed information can be found in [13].

Architecture of the Virtual Rap Dancer

The (real time) architecture of the virtual rap dancer consists of several main parts. The sensor channels analyze incoming audio and video in order to detect beats in each separate channel to which the dancer can time its dance moves. The beat predictor module combines the different streams of detected beats, trying to merge beats that were detected in two different channels at the same time into one beat and trying to predict when a next beat is most likely to occur. This prediction is then used by the motion controller that will plan a next dance move in such a way that its focus point will coincide with the predicted next beat. Input obtained from the dance mat allows the system to detect whether the person is dancing with his/her feet. This information is used to choose appropriate dance moves from the lexicon. The animation system finally will execute the planned movements after adapting them to some style parameters.

Analyzing and combining input

There are two types of modules that attempt to recognize beats. The audio system does this largely based on the energy of the audio signal. This signal can e.g. come from a (rap) song or from vocal percussion. The video system tracks the face and hands of the person in the camera view (based on [12]), and recognizes beats based on the hands or face crossing (implicit) trigger-lines in the image. Two incoming beats that are too close together are assumed to be the same beat recognized by different sources. A simple beat-prediction algorithm takes the time-span between the previous beat (from any source) and this beat, and uses that to predict when a next beat is expected. We are currently working on an input module that extracts more advanced characteristics of the human dancer from the video input, such as the *type* of moves used by the human dancer, or the *style of dancing*.

Generating dance moves from input and style parameters

The animation module uses the database of rap-dance moves which was constructed from the analysis of example dancers. Moves are stored in a database as joint angles for key frames, plus information such as which key frames are to be performed on a (musical) beat. The movement controller selects moves to be executed based on some style parameters. These moves are then planned using the information about which key frames are to be aligned to musical beats and the beat prediction. Finally the dance moves are modified with a style-adaptation. We are currently working on an animation system that allows more freedom in the style of execution of dance moves.

Conclusion

In Figure 2 we show the set-up of the system in a demonstration booth. Visitors are invited to start dancing, their dancing movements are tracked and

guide the virtual rap dancer in its own dance movements and gestures. There is input from a camera, a dance mat and the music that is played in the environment. In the near future we continue developing the interactivity between human and virtual dancer. In particular we will spend more efforts on developing rapport between human and virtual dancer.

Figure 2. Demonstration set-up

Acknowledgements

The first version of the virtual rap dancer based on analyzed video clip was made by our students. Interaction through microphone, camera and dance mat was realized with the help of Dennis Hofs and Joost Vromen. A more complete description of an earlier version of the virtual rap dancer can be found in [13].

Citations

[1] Tosa, N., Nakatsu, R. Emotion recognition-based interactive theatre –Romeo & Juliet in Hades -. *Eurographics '99*, M.A. Alberti, G. Gallo & I. Jelinek (Eds.) (1999)

[2] Pinhanez, C., Bobick, A. Using computer vision to control a reactive graphics character in a theater play. *Proc ICVS '99* (1999)

[3] Bertolo, M., Maninetti, P., Marini, D. Baroque dance animation with virtual dancers. *Eurographics '99*, M.A. Alberti, G. Gallo & I. Jelinek (Eds.) (1999)

[4] Kragtwijk, M., Nijholt, A., Zwiers, J. An animated virtual drummer. *Int. Conf. Augmented, Virtual Environments and Three-dimensional Imaging (ICAV3D)*, V. Giagourta and M.G. Strintzis (eds.), Mykonos, Greece (2001) 319-322

[5] Goto, M., Muraoka, Y. A Virtual Dancer "Cindy" Interactive Performance of a Music-controlled CG Dancer, *Proc. Lifelike Computer Characters* (1996) 65

[6] Hamanaka, M., Goto, M., Asoh, H., Otsu, N. A learning-based jam session system that imitates a player's personality model. *Proc. IJCAI (2003)* 51-58

[7] Friberg, A., Schoonderwaldt, E., Juslin, P.N., Bresin, R. Automatic real-time extraction of musical expression. *Int. Comp. Music Conf. (2002)* 365-367

[8] Mancini, M., Bresin, R., Pelachaud, C. From acoustic cues to expressive ECAs. *6th Int. Wsh. on Gesture in HCI and Simulation.* (2005)

[9] Taylor, R., Torres, D., & Boulanger, P. Using music to interact with a virtual character. *Int. Conf. on New Interfaces for Musical Expression* (2005) 220-223

[10] Taylor, R., Boulanger, P., Torres, D. Visualizing emotion in musical performance using a virtual character. *5th Int. Symp. on Smart Graphics* (2005)

[11] Camurri, A., Lagerlöf, I., Volpe, G. Recognizing Emotion from Dance Movement: Comparison of Spectator Recognition and Automated Techniques. *Int. J. of Human-Computer Studies*, 59(1-2) (2003) 213-225

[12] Poppe, R., Heylen, D., Nijholt, A., Poel, M. Towards real-time body pose estimation for presenters in meeting environments. *Int. Conf. in Central Europe on CG, Visualization and Computer Vision* (2005)

[13] Reidsma, D., Nijholt, A., Rienks, R. & Hondorp, H. Interacting with a Virtual Rap Dancer. *Proc. INTETAIN'05*, LNAI 3814, 132-141.