Gesture Capture: Paradigms in Interactive Music/ Dance Systems

FRÉDÉRIC BEVILACQUA, NORBERT SCHNELL, SARAH FDILI ALAOUI

INTRODUCTION

Electronic or digital interactive systems have been experimented with in dance for more than fifty years. The piece *Variations V* by Merce Cunningham and John Cage, performed in 1965 with dancers interacting with analog electronic sound systems, is one such groundbreaking case (cf. Miller 2001). The use of interactive digital technology grew in importance in the 1990s with the advent of affordable sensor technologies and software dedicated to motion tracking, such as *Eycon*, *EyesWeb*, *BigEye*, and *softVNS*, which were developed for artistic use. The first decade of the 21st century showed a steady increase of new experimentation and usage of media technologies in various dance contexts and aesthetics, including performances and installations.

This text gives some idea of the wide range of currently available technological tools used to sense gesture and movement. Most importantly, we would like to emphasize that the current discourse on interactive systems has moved away from 'experimenting' with technology, which is nowadays ubiquitous, to more fundamental questions on the description and notation of gesture and movement (cf. deLahunta/Bevilacqua 2007), and what transmission these systems could provide or facilitate. Several choreographers and dance companies have built ambitious interdisciplinary research projects (cf. InsideMovement Knowledge.net; SynchronousObjects.osu.edu) involved with such questions. These initiatives reflect the converging interests of different disciplines – dance, music, engineering and cognitive sciences – towards gesture research. For example, research on sensorimotor learning has influenced the Human Computer Interaction field, where the role of action and gesture has increased significantly (cf. Dourisch 2001; Leman 2007).

Working in such an interdisciplinary context at Ircam (Institut de Recherche et Coordination Acoustique/Musique)[1], we have developed, in collaboration with choreographers/composers/media artists, computer based gesture analysis and interactive audio processing systems that allows performers to control or interact with digital media – sound or video (cf. Bevilacqua 2007). At the Dance Congress 2009, we presented a lecture/demonstration of these tools and explained paradigms that are central to these applications. Summarizing key elements of our presentation here, we will first categorize the different sensing systems typically found in dance contexts, in order to clarify what the term ‘gesture capture’ can encompass. In the second part, we will provide examples of gesture sound controls often found in interactive dance performances.

SENSING FOR INTERACTING

It is possible to argue that in any interactive dance system, the technical constraints related to the chosen gesture capture apparatus influences the choreographic work. Therefore, it is generally useful to describe technical constraints that might inform some aspects of the work. Nevertheless, we wish to emphasize here the interaction paradigms that are associated with gesture technology instead of simply describing the technical features of specific systems. For this reason, we propose to classify the different approaches for sensing gesture as used in dance performances or installations using three main categories, ‘body’, ‘space’, and ‘time’. This classification helps to clarify the metaphors implicitly or explicitly related to interactive systems. Obviously, these categories should not be considered absolute, or their limits as definite: most interactive systems will generally include more than one of these categories.

---

1 Ircam is one of the largest public research centers dedicated to both musical expression and scientific research. This article relates specifically to research performed by The Real Time Musical Interaction and the Performing Arts Technology Research teams, which have been collaborating regularly with various choreographers since 2003, developing interactive systems for the performing arts, such as the *gesture follower* and audio processing tools.

Body

Although it is clear that physical movement is largely used in interactive systems, we first also want to call to mind that numerous other types of body-centered interaction are possible by using, for example, physiological signals.

Physiological signals

All kinds of Measurements of physiological parameters can be utilized in interactive systems. Technically these systems are generally adapted from technology developed for biofeedback. Mechanisms such as muscle activation, for example, can be measured in form of electrical activity with sensors that are put in contact with the skin (electromyography). Such systems have been incorporated in dance performances (cf. Palindrome.de) or music performance (cf. Tanaka/Knapp 2002). Interestingly, these techniques can also be sensitive to 'pre-movements' or muscle tension even if there is no significant visible movement.

Other types of physiological quantities have also been used in media performance including breathing[2], the heartbeat, and even brain waves. Skin contact between two people, or between the skin and an (electrically conductive) object can also be easily measured thanks to the skin's electrical conductivity. This property allows for the design of 'touch sensitive' objects.[3] Performances have also been designed that take advantage of this effect, sonifying skin contact between performers and even the public (cf. Woudi-Tat.org).

Body posture and movement

Different types of technological systems enable the measurement of body posture and motion. First, sensors can be attached to the body as illustrated in Figure 1. Miniature accelerometers and gyroscopes are, for example, sensitive to inclination, rotation and acceleration (note: these later types of sensors are now found

2 Myriam Gourfink, for example, used a breathing sensor in her piece *This is my House*, along with other motion sensors, http://myriam-gourfink.com/thisIsMyHouse.htm (April 5, 2010).

3 The crackle box invented by Michel Waisvisz is an historic musical example, http://www.crackle.org (April 5, 2010).

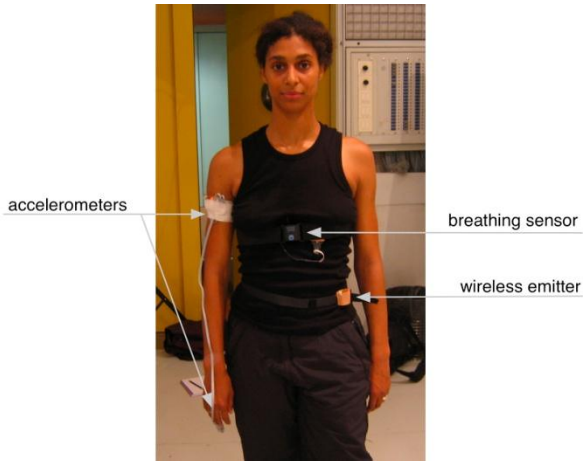
for example in gaming interfaces such as the Wiimote); bending sensors measure joint angles.[4] Generally, the sensors are connected to wireless data emitters.

Second, video systems are efficient in capturing postures and movements of the entire body remotely. For example, a single camera system can track the dancer's silhouette, as shown in Figure 2, taken from the *Double Skin/Double Mind* installation (cf. InsideMovementKnowledge.net).

A large number of computer programs characterize body movement by tracking the whole silhouette or specific body parts. The Eyseweb software, for example, computes parameters such as 'quantity of motion', 'direction', and 'stability' among several others (cf. Camurri/Mazzarino/Volpe 2004). This software was recently used to automatically annotate short videos of a web dance database (cf. Tardieu/Chessini et al. 2009).

In single camera systems, the measurement of the movement highly depends on the position of the dancer relative to the camera. To avoid such problems, multiple cameras are required. Typically, 3D optical motion-capture systems, initially developed for biomechanic studies or for animation, allow for the 3D reconstruction of a simplified skeleton. Nevertheless, they require the use of small reflective markers on the body and are generally complex to handle in real-time and in performance situations. Dance performances using such systems have been relatively rare, with few notable exceptions such as two pieces by Bill T. Jones and Trisha Brown, developed at the Arizona State University, in collaboration with the OpenEndedGroup (cf. Downie 2005).

Figure 1: Sensors for breathing and acceleration measurements



4 See the mini dancer, http://www.troikaranch.org (April 5, 2010).

Figure 2: Video tracking of the dancer silhouette and analysis of the size of different parts of the silhouette (from the Double Skin/Double Mind *installation).*

Photography: Thomas Lenden

Space

We refer here to paradigms where some properties of the space are explicitly included in the interaction. This can, for example, imply defining particular zones of the space, in which the presence of the user triggers specific electronic events. This type of paradigm is among one of the first implemented historically, using either light barriers or camera systems. In the experimental piece *Variations V* mentioned earlier, proximity to sensors (Theremin antenna) were placed on particular spots on stage and reacted to dancers approaching these particular spots.

Generally, space-based interaction implies structuring the space, and associating audio/video processes with specific spatial location. Commons paradigms are, for example: body presence/absence, crossing borders, entering/leaving zones. Obviously, motion can also be naturally associated with these interactions, for example, by measuring the 'quantity of motion' in a particular spatial zone. Nevertheless, we would like to point out that in these cases the motion remains referenced to absolute spatial locations, and not relative to the body itself as described in the previous section.

Time

At first, it might seem unclear how 'time' can be referred to a category of interaction. We argue that 'temporal interaction' can be put forward in a similar fashion as spatial interaction (cf. Bevilacqua 2007). Similar to spatial limits or zones, one can define time limits and time moments. Moreover, interaction can be based on synchronizing specific dance gestures and sound/visual processes. In other words, interaction can be driven by temporal events, time sequences and synchronization mechanisms.

Generally, this interaction paradigm relies on software that is designed to analyze temporal data, i.e. a sequence of events or postures. Our research currently aims at taking into account gesture data as 'time processes'. This implies considering basic elements such as 'phrases' and 'transitions' (as opposed to 'postures') in relationship to time-based media such as sound or video (we will describe in more detail possible interaction models in the next section). Note that this approach was motivated in part by collaboration with choreographers, who pointed out the necessity of considering gestures as continuous time-related processes.

EXAMPLES OF GESTURE CONTROLLED SOUND INTERACTION

Digital sound processes can be controlled or altered using the different types of gesture parameters we have described in the previous section. Establishing the relationship between gesture parameters and the actual sound properties is a central task when building any interactive system. Such a procedure is often referred as a gesture-to-sound "mapping" (cf. Wanderley 2002; Bevilacqua/Muller/Schnell 2005).

In the following, we will describe concrete examples ranging from relatively simple mappings to more elaborate scenarios working with complex gestures analysis and audio-visual processes. In the first two sections below, we will define simple relationships between gesture and sound rendering, corresponding to explicit interaction metaphors. By defining this relationship, we can create a sort of 'musical instrument' that can be 'played' by the dancer. In the last two sections, we will introduce the possibility of handling complex phrases in the interaction design, which can lead to more abstract relationships between gesture and sound.

Triggering sound events

Triggering is one of the most simple and common processes used in interactive systems. As an introductory example, we can show that ‘percussive’ gesture (i.e. strokes measured with ‘accelerometers’) can directly trigger percussive sound events. The gesture ‘intensity’ can furthermore affect the volume and characteristics of the sound.

An application of this paradigm is the sequential triggering of discrete recorded sound events. For example: the dancer selects, in a preliminary step, particular sound events in a recording. Then, each sound event can be played one by one by the dancer using percussive gestures. The dancer can also control the tempo and rhythm of the sequence of the recorded sound. Experiments generally show that the clarity of this interaction allows for a rapid appropriation of the sound control by the dance movements. Nevertheless, this influences performers towards performing discrete strokes or accents. Continuous sound control appears then as a natural extension, as explained in the next section.

Continuous control of sound grains

Continuous movement parameters, e.g. inclination or velocity, can naturally be ‘mapped’ to continuous sound parameters. Examples based on granular synthesis techniques, which have been widely used in dance/music systems, were among the many techniques that we demonstrated during the workshop. Granular synthesis is based on segmenting sound recordings in small ‘sound grains’ and then playing them in such as way as to create sound textures. The original sound characteristics can be either preserved or radically altered.

We experimented in particular with sound recordings related to natural, human or mechanical movements, sounds of liquid pouring, rolling/rotating objects, human beat boxing and machines. These sounds can be easily decomposed into very short elementary elements (i.e. sound grains) and recomposed according to gestural input (cf. Schnell/Borghesi et al. 2005; Schnell/Röbel et al. 2009). In simple cases, dynamic movement parameters such as ‘energy’ can be directly used to control the intensity of rendered sound textures.

More complex relationships make use of an intermediate model mediating specific behaviors between gesture input and sound responses. A compelling example is based on the rainstick metaphor.[5] In detail, the sound rendering can simulate the sound of various materials (water, stones or abstract sound textures) as if agitated and moved from one side to the other of an object, according to the inclination of the object. By holding a real object containing an inclination sensor, dancers therefore control the sound of various virtual 'sound grains' pouring from one side to the other side of the object. By directly holding the inclinometer sensor to their body, dancers can even directly embody the sound object, 'pouring sound grains' by bending their body.

Gesture recognition for media control

Gesture recognition systems are particularly useful where interpretation of sensor parameters becomes complex, leading to cumbersome programming. Even simple gestures and body movements may in fact generate a very large number of movement parameters and complex data patterns.

Using a gesture recognition system can simplify the setting of the interaction and offers possibilities of using longer choreographed movements. A first step is to define 'phrases', i.e. gestures units, that the computer system must learn in a first phase in order to be able to recognize them automatically in a second phase. Interestingly, this approach lets the dancer define a gesture 'vocabulary' and thus work on a symbolic level. Over the past years, we have developed a system at Ircam called the *gesture follower* that can be used for gesture recognition of complex postures, phrases or trajectories (cf. Bevilacqua/Guédy et al. 2007; Bevilacqua/Zamborlin et al. 2009). The *gesture follower* has been used in dance performances (cf. thebakery.org), interactive installations (cf. if-then-installed.leprojet.net) and in music pedagogy (cf. Bevilacqua/Guédy et al. 2007). To use the *gesture follower*, the dancer first records phrases, using sensors or video systems, to define a vocabulary. The control of audio processes (triggering, synchronization and continuous sound control) can then be built on the basis of this vocabulary.

In the case of the *Double Skin/Double Mind* installation motion parameters were fed into the analyzing system, which was in this case especially tuned to movements principles defined by the Emio Greco | PC company. The results of the analysis could either be connected to sounds or visual feedback.

5 Such a paradigm was used in the installation *Grainstick* by Pierre Jodlowski, see: http://agora2010.ircam.fr/935.html?event=887&L=1 (April 5, 2010).

In the case of a collaboration with Richard Siegal, we developed another application of the *gesture follower* for an installation (if-then-installed.leprojet.net) and a multimedia dance performance. In *Homo Ludens* (cf. thebakery.org), Richard Siegal improvises at the beginning of the piece with a set of dance phrases that are recognized in real-time by the systems (he wore motion sensors on his wrists). When recognized, pre-recorded videos of theses phrases were displayed, creating a sort of dialog between the dancer and the videos.

Beyond recognizing phrases, the *gesture follower* allows for the synchronization of arbitrary gestures and movements to time-based media, such as audio and video recordings. Once a particular dance phrase and recording have been entered in the system, it can control in real-time the synchronized rendering of the recording according to the gesture variations. More precisely, the system can continuously control the pace and mix of digital media (rather than just triggering start/stop). In other words, the dancer can continuously control the choices and the temporality (i.e. tempo, rhythm, order) of recording rendered by her/his performance. Therefore, the interaction paradigm enabled by the *gesture follower* equals intrinsically translating the 'temporal' unfolding of gestures to the 'temporal' unfolding of digital media.

CONCLUSION

In this paper we have given examples of sensing techniques for dance-music interactive systems. We proposed to categorize the different paradigms as 'body'-, 'space'- or 'time'-related. The combination of these different paradigms can lead to different layers of computer-mediated interaction between dance and sound/visual processes. The combination of both simple interaction paradigms with recent advances on gesture recognition and following currently gives rise to the novel experiments that we are pursuing. Important challenges lie now in the use of interactive systems with coherent gesture descriptions that could be shared by dancers, musicians and engineers. Recent productions (e.g. *If/Then Installed*) and research projects (i.e. *Inside Movement Knowledge*) that we have participated in are very promising in this regard. Furthermore, we will continue to pursue research on notions such as 'quality of movements' that could be derived from gesture capture system. We believe that such analysis should further enrich interaction paradigms with new media.

We acknowledge partial support of the following projects: *EarToy* and *Interlude* (ANR – French National Research Agency) and *SAME* (EU – ICT). We thank all the members of the Real-Time Musical Interactions team at Ircam.

REFERENCES

Bevilacqua, Frédéric/Muller, Rémy/Schnell, Norbert (2005): “Mnm: a max/msp mapping toolbox”, in: NIME 05: Proceedings of The 2005 International Conference on New Interfaces for Musical Expression, Vancouver, pp. 85–88.

Bevilacqua, Frédéric (2007): “Momentary Notes on Capturing Gestures”, in: Emio Greco | PC and the Amsterdam School of the Arts (eds.), (Capturing Intentions), Amsterdam: Emio Greco | PC and the Amsterdam School of the Arts.

Bevilacqua, Frédéric/Guédy, Fabrice/Schnell, Norbert et al. (2007): “Wireless sensor interface and gesture-follower for music pedagogy”, in: NIME 07: Proceedings of The Seventh International Conference on New Interfaces for Musical Expression, New York, pp. 124–129.

Bevilacqua, Frédéric/Zamborlin, Bruno/Sypniewski, Anthony et al. (2009): “Continuous realtime gesture following and recognition”, in: Lecture Notes in Computer Science Volume 5934, Berlin/Heidelberg: Springer, pp. 73-84.

Camurri, Antonio/Mazzarino, Barbara/Volpe, Gaultiero (2004): “Analysis of expressive gesture: The eyesweb expressive gesture processing library”, in: Antonio Camurri/Gaultiero Volpe (eds.), Gesture-Based Communication in Human-Computer Interaction (= LNAI 2915), Berlin [a.o.]: Springer, pp. 460-467.

deLahunta, Scott/ Bevilacqua, Frédéric (2007): “Sharing Descriptions of Movement”, in: International Journal of Performance and Digital Media 3/1, pp. 3-16.

Dourish, Paul (2001): Where The Action Is: The Foundations of Embodied Interaction, Cambridge, MA: MIT Press.

Downie, Marc (2005): Choreographing the Digital Agent: Live Performance Graphics for Dance Theater, Cambridge, MA: MIT PhD thesis.

Leman, Marc (2007): Embodied Music Cognition and Mediation Technology, Cambridge, MA: MIT Press.

Miller, Leta E. (2001): “Cage, Cunningham, and Collaborators: The Odyssey of Variations V.”, in: The Music Quarterly Volume 85, Number 3, pp. 545-567.

Emio Greco | PC and the Amsterdam School of Arts (eds.) (2007): (Capturing Intentions), Amsterdam: Emio Greco | PC and the Amsterdam School of Arts.

Schnell, Norbert/Borghesi, Riccardo/Schwarz, Diemo et al. (2005): "Ftm – complex data structures for max.", proceedings of the International Computer Music Conference (ICMC), Barcelona, Spain, September 2005.

Schnell, Norbert/ Röbel, Axel/ Schwarz, Diemo et al. (2009): Assembling tools for content based real-time interactive audio processing in max/msp, proceedings of the International Computer Music Conference (ICMC), Montreal, Canada, August 2009.

Tardieu, Damien/Chessini, Ricardo/ Dubois, Julien et al. (2009): Video navigation tool: Application to browsing a database of dancers' performances, QPSR of the numediart research program, Volume 2, Number 3, September 2009.

Tanaka, Atau/Knapp, Benjamin R.: "Multimodal interaction in music using the Electromyogram and relative position sensing", in: Proceedings of the 2002 Conference on New Interfaces for Musical Expression, Dublin, Ireland, May 24-26, 2002, pp. 43-48.

Wanderley, Marcelo (ed.) (2002): Mapping Strategies in Real-Time Computer Music (= Organised Sound 7/02), Cambridge: Cambrigde University Press.

WEBSITES

Agora 2010: http://agora2010.ircam.fr/935.html?event=887&L=1 (April 5, 2010).

Crackle: http://www.crackle.org/ (April 5, 2010).

If Then Installed: http://if-then-installed.leprojet.net/ (April, 5 2010).

Inside Movement Knowledge: http://insidemovementknowledge.net/ (April 5, 2010).

Myriam Gourfink: http://www.myriam-gourfink.com/thisIsMyHouse.htm (April 5, 2010).

Palindrome Inter.media Performance Group: http://www.palindrome.de/ (April 5, 2010).

Synchronous Objects: http://synchronousobjects.osu.edu/ (April 5, 2010).

The Bakery: http://www.thebakery.org/index02.html (April 5, 2010).

The OpenEnded Group: http://openendedgroup.com/ (April 5, 2010).

Troika Ranch: http://www.troikaranch.org (April 5, 2010).

Woudi-Tat: http://www.woudi-tat.org/ (April 5, 2010).