EURASIP Journal on Advances in Signal Processing

Dance-the-Music: an educational platform for the modeling, recognition and audiovisual monitoring of dance steps using spatiotemporal motion templates

EURASIP Journal on Advances in Signal Processing 2012,
2012:35 doi:10.1186/1687-6180-2012-35

Pieter-Jan Maes (maes.pieterjan@gmail.com)
Denis Amelynck (denis.amelynck@UGent.be)
Marc Leman (marc.leman@UGent.be)

ISSN	1687-6180
Article type	Research
Submission date	15 April 2011
Acceptance date	16 February 2012
Publication date	16 February 2012
Article URL	http://asp.eurasipjournals.com/content/2012/1/35

Dance-the-Music: an educational platform for the modeling, recognition and audiovisual monitoring of dance steps using spatiotemporal motion templates

Pieter-Jan Maes*, Denis Amelynck and Marc Leman

IPEM, Department of Musicology, Ghent University, Blandijnberg 2, 9000 Ghent, Belgium

*Corresponding author: pieterjan.maes@UGent.be

Email address:

DA: denis.amelynck@UGent.be

ML: marc.leman@UGent.be

Abstract

In this article, a computational platform is presented, entitled "Dance-the-Music", that can be used in a dance educational context to explore and learn the basics of dance steps. By introducing a method based on spatiotemporal motion templates, the platform facilitates to train *basic step models* from sequentially repeated dance figures performed by a dance teacher. Movements are captured with an optical motion capture system. The teachers' models can be visualized from a first-person perspective to instruct students how to perform the specific dance steps in the correct manner. Moreover, recognition algorithms-based on a template matching method-can determine the quality of a student's performance in real time by means of multimodal monitoring techniques. The results of an evaluation study suggest that the Dance-the-Music is effective in helping dance students to master the basics of dance figures.

Keywords: dance education; spatiotemporal template; dance modeling and recognition; multimodal monitoring;

audiovisual dance performance database; dance-based music querying and retrieval.

1 Introduction

Through dancing, people encode their understanding of the music into body movement. Research has shown that this body engagement has a component of temporal synchronization but also becomes overt in the spatial deployment of dance figures [1–5]. Through dancing, dancers establish specific spatiotemporal patterns (i.e., dance figures) in synchrony with the music. Moreover, as Brown [1] points out, dances are modular in organization, meaning that the complex spatiotemporal patterns can be segmented into smaller units, called gestures [6]. The beat pattern presented in the music functions thereby as an elementary structuring element. As such, an important aspect of learning to dance is learning how to perform these basic gestures in response to the music and how to combine these gestures to further develop complex dance sequences.

The aim of this article is to introduce a computational platform, entitled "Dance-the-Music", that can be used in dance education to explore and learn the basics of dance figures. A special focus thereby lays on the spatial deployment of dance gestures, like footstep displacement patterns, body rotation, etc. The platform facilitates to train *basic step models* from sequentially repeated dance figures performed by a dance teacher. The models can be stored together with the corresponding music in audiovisual databases. The contents of these databases, the teachers' models, are then used (1) to give instructions to dance novices on how to perform the specific dance gestures (cf., dynamic dance notation), and (2) to recognize the quality of students' performances in relation to the teachers' models. The Dance-the-Music was designed explicitly from a user-centered perspective, meaning that we took into account aspects of human perception and action learning. Four important aspects are briefly described in the following paragraphs together with the technologies we developed to put these aspects into practice.

Spatiotemporal approach When considering dance gestures, time-space dependencies are core aspects. This implies that the spatial deployment of body parts is directly linked to the temporal structure outlined in the music (involving rhythm and timing). The modeling and automatic recognition of dance gestures often involve Hidden Markov Modeling (HMM) [7–10]. However, HMM has the property to exhibit some degree of invariance to local warping (compression and stretching) of the time-axis [11]. Even though this might be an advantage for applications like speech recognition, it is a serious drawback when considering spatiotemporal relationships in dance gestures. HMMs are fine for detecting basic steps and spatial patterns but cause major difficulties for timing aspects because of the inherent time-warping mechanism. Therefore, for the Dance-the-Music, we will introduce an approach based on spatiotemporal motion templates [12–14]. As will be explained in depth, the discrete time signals representing the gestural parameters extracted from dance movements are organized into a fixed-size multidimensional feature array forming the spatiotemporal template. Dance gesture recognition will be achieved by a template matching technique based on cross-correlation computation.

User- and body-centered approach The Dance-the-Music facilitates to instruct dance gestures to dance novices with the help of an interactive visual monitoring aid (see Sections 3.4.1 and 4). Concerning the visualization of basic step models, we take into account two aspects involving the perception and understanding of complex multimodal events, like dance figures. First, research has shown that segmentation of ongoing activity into smaller units is an automatic component of human perception and functional for memory and learning processes [1,15]. For this, we applied algorithms that segment the continuous stream of motion information into a concatenation of elementary gestures (i.e., dance steps) matching the beat pattern in the music (cf., [6]). Each of these gestures is conceived as a separate unit, having a fixed start- and endpoint. Second, neurological findings indicate that motor representations based on first-person perspective action involve, in relation to a third-person perspective, more kinesthetic components and take less time to initiate the same movement in the observer [16]. Although applications in the field of dance gaming and education often enable a manual adaptation of the viewpoint perspective, they do not follow automatically when users rotate their body during dance activity [17–20]. In contrast, the visual monitoring aid of the Dance-the-Music automatically adapts the viewpoint perspective in function of the rotation of the user at any moment.

Direct, multimodal feedback The most commonly used method in current dance education to instruct dance skills is the *demonstration-performance method*. As will explained in Section 2, the Dance-the-Music elaborates on this method in the domain of human-computer interaction (HCI) design. In the demonstration-performance method, a model performance is shown by a teacher which must then be imitated by the student under close supervision. As Hoppe et al. [21] point out, a drawback to this learning schematic is the lack of an immediate feedback indicating how well students use their motor apparatus in response to the music to produce the requisite dance steps. Studies have proven the effectiveness of self-monitoring through audiovisual feedback in the process of acquiring dancing and other motor skills [19, 22–24]. The Dance-the-Music takes this into account and provides direct, multimodal feedback services. It is in this context that the recognition algorithms—based on template matching—have their functionality (see Section 3.3). Based on cross-correlation computation, they indicate how well a student's performance of a specific dance figure matches the corresponding model of the teacher.

Dynamic, user-oriented framework The Dance-the-Music is designed explicitly as a computational framework (i.e., a set of algorithms) of which content and configuration settings are entirely dependent on the needs and wishes of the dance teacher and student. The content mainly consists of the dance figures that the teacher wants to instruct to the student and the music that corresponds with it. Configuration settings involve tempo adjustment, the number of steps in one dance figure, the number of cycles to perform to train a model, etc. Moreover, the Dance-the-Music is not limited to the gestural parameters presented in this article. Basic programming skills facilitate to input data of other motion tracking/sensing devices, extract other features (acceleration, rotational data of other body parts, etc.), and add these into the model templates. This flexibility is an aspect that distinguishes the Dance-the-Music from commercial hardware (e.g., dance dance revolution [DDR] dancing pad interfaces) and software products (e.g., StepMania for Windows, Mac, Linux; DDR Hottest Party 3 for Nintendo Wii; DanceDanceRevolution for PlayStation 3, DDR Universe 3 for Xbox360, Dance Central and Dance Evolution for Kinect, etc.). Most of these systems use a fixed, built-in vocabulary of dance moves and music. Another major downside to most of these commercial products is that they provide only a small action space restricting spatial displacement, rotation, etc. The Dance-the-Music drastically expands the action/dance space facilitating rotation, spatial displacement, etc.

The structure of the article is as follows: In Section 2, detailed information is provided about the methodological grounds on which the instruction method of the educational platform is based. Section 3 is then dedicated to an in-depth description of the technological, computational, and statistical aspects underlying the design of the Dance-the-Music application. In Section 4, we present a user study conducted to evaluate if the system can help dance novices in learning the basics of specific dance steps. To conclude, we discuss in Section 5 the technological and conceptual performance and future perspectives of the application.

2 Instruction method

In concept, the Dance-the-Music brings the traditional demonstration-performance approach into the domain of HCI design (see Section 1). Although the basic procedure of this method (i.e., teacher's demonstration, student's performance, evaluation) stays untouched, the integration of motion capture and real-time computer processing drastically increase possibilities. In what comes, we outline the didactical procedure incorporated by the Dance-the-Music in combination with the technology developed to put it into practice.

2.1 Demonstration mode

A first mode facilitates dance teachers to train basic step models from their own performance of specific dance figures. Before the actual recording, the teacher is able to configure some basic settings, like the music on which to perform, the tempo of the music, the number of steps per dance figure, the amount of training cycles, etc. (see module 1 and 2, Figure 1). Then, the teacher can record a sequence of a repetitive performed dance figure of which the motion data is captured with optical motion capture technology (see module 3, Figure 1). When the recording is finished, the system immediately infers a basic step model from the recorded training data. The model can then be displayed (module 4, Figure 1) and, when approved, stored in a database together with the corresponding music (module 5, Figure 1). This process can then be repeated to create a larger audiovisual database. These databases can be saved as .txt files and loaded whenever needed.

2.2 Learning (performance) mode

By means of a visual monitoring aid (see Figure 2, left) with which a student can interact, the teachers' models can be graphically displayed from a first-person perspective and can be segmented into individual steps. By imitating the graphically notated displacement and rotation patterns, a dance student learns how to perform the step patterns in a proper manner. In order to support the dance novice, the playback speed of the dynamic visualization is made variable. When played in the original tempo, the model can be displayed in synchrony with the music that corresponds with it. Moreover, recognition algorithms are implemented facilitating a comparison between the model and the performance of the dance novice (see Section 3.3). As such, direct multimodal feedback can be given monitoring the quality of a performance (see Section 3.4).

2.3 Gaming (evaluation) mode

Once students learned to perform the dance figures with the visual monitoring aid, they can exhibit their dance skills. This is the application mode allowing students to literally "Dance the Music". By performing a specific dance figure learned with the visual monitoring aid, students receive music that fits a particular dance genre. It is in this context of gesture-based music retrieval that the recognition algorithms based on template matching come to the fore (see Section 3.3). Based on cross-correlation computation, these algorithms detect how exact a performed dance figure of a student matches the model performed by the teacher. The quality of the student's performance in relation to the teacher's model is then expressed in the auditory feedback and in a numerical score stimulating the student to improve his/her performance.

The computational platform itself is built in Max/MSP (www.cycling74.com). The graphical user interface (GUI) can be seen in Figure 1. It can be shown on a normal computer screen or projected on a big screen or on the ground. One can interact with the GUI with a computer mouse. The design of the GUI is kept simple to allow intuitive and user-friendly accessibility.

3 Technical design

Different methods are used for modeling and recognizing movement (e.g., HMM-based, template-based, state-based, etc.). For the Dance-the-Music, we have made the deliberate choice to implement a template-based approach to gesture modeling and recognition. In this approach, the discrete time signals representing the gestural parameters extracted from dance movements are organized into a fixed-size multidimensional feature array forming the spatiotemporal template. For the recognition of gestures, we will apply a template matching technique based on cross-correlation computation. A basic assumption in this method is that gestures must be periodic and have similar temporal relationships [25, 26]. At first sight, HMMs or dynamic time warping (DTW)-based approaches might be understood as proper candidates. They facilitate learning from very few training samples (e.g., [27, 28]) and a small number of parameters (e.g., [29]). However, HMM and DTW-based methods exhibit some degree of invariance to local time-warping [11]. For dance gestures in which rhythm and timing are very important, this is problematic. Therefore, when explicitly taking into account the spatiotemporal relationship of dance gestures, the template-based method we introduce in this article provides us with a proper alternative.

In the following sections, we first go into more detail how dance movements are captured (Section 3.1). Afterwards, we will explain how the raw data is pre-processed to obtain gestural parameters which are expressed explicitly from a body-centered perspective (Section 3.1.2). Next, we will point out how the Dance-the-Music models (Section 3.2) and automatically recognizes (Section 3.3) performed dance figures using spatiotemporal templates and how the system provides audiovisual feedback of this performance (Section 3.4). A schematic overview of Section 3 is given in Figure 3.

3.1 Motion capture and pre-processing of movement parameters

Motion capture is done with an infrared (IR) optical system (OptiTrack/Natural Point). Because we are interested in the movements of the body-center and feet, we attach *rigid bodies* to these body parts (see Figure 4). The body-center (i.e., center-of-mass) of a human body in standing position is situated in the pelvic area (i.e., roughly the area in between the hips). Because visual occlusion can occur (with resulting data loss) when the hands cover hip markers, it can be opted to attach them to the back of users instead (see Section 3.1.2, par. Spatial displacement). A rigid body consists of minimum three IR-reflecting markers of which the mutual distance is fixed. As such, based on this geometric relationship, the motion capture system is able to identify the different rigid bodies. Furthermore, the system facilitates to output (1) the 3-D position of the centroid of a rigid body, and (2) the 3-D rotation of the plane formed by the three (or more) markers. Both the position and rotation components are expressed in reference to a global coordinate system predefined in the motion capture space (see Figure 5). These components will be referred to as *absolute*, in contrast to their relative estimates in reference to the body (see Section 3.1.1).

For the Dance-the-Music, the absolute (x, y, z) values of the feet and body-center together with the rotation of the body-center expressed in quaternion values (q_x, q_y, q_z, q_w) are streamed, using the open sound control (OSC) protocol to Max/MSP at a sample rate of 100 Hz.

3.1.1 Relative position calculation

The position and rotation values of the rigid body defined at the body-center are used to transform the absolute position coordinates into relative ones in reference to a body-fixed coordinate system with an origin positioned at the body-center (i.e., local coordinate system). The position and orientation of that local coordinate system in relation to the person's body can be seen in more detail in Figure 5. The transformation from the initial body stance (Figure 5, left) is executed in two steps. Both are incorporated in real-time operating algorithms, implemented in Max/MSP as java-coded *mxj*-objects.

1. Rotation of the local, body-fixed coordinate system in a way it has the same orientation as the global coordinate system (Figure 5, middle). What actually happens, is that all absolute (x, y, z) values are rotated based on the quaternion values of the rigid body attached to the body-center representing the difference in orientation between the local and the global coordinate system.

2. Displacement of the origin (i.e., body-center) of the local, body-fixed coordinate system to the origin of the global coordinate system (Figure 5, right).

As such, all position values can now be interpreted in reference to a person's own body-center. However, a problem inherent to this operation is that rotations of the rigid body attached to the body-center, independent from actual movement of the feet, do result in apparent movement of the feet. The consequences for free movement (for example for the upper body) are minimal when taking into account a well-considered placement of the rigid body attached to the body-center. The placement of the rigid body at the hips, as shown in Figure 4, does not constrain three-dimensional rotations of the upper body. However, the problem remains for particular movements in which rotations of the body-center other than the rotation around the vertical axis are important features, like lying down, rolling over the ground, movements where the body-weight is (partly) supported by the hands, flips, etc. Apart from the problems they cause for the mathematical procedures presented in this section, these movements are also incompatible with the visualization strategy which is discussed into more detail in Section 3.4.1. As such, these movements are out of the scope of the Dance-the-Music.

3.1.2 Pre-processing of movement parameters

As already mentioned in the introduction, the first step in the processing of the movement data is to segment the movement performance into discrete gestural units (i.e., dance steps). The borders of these units coincide with the beats contained in the music. Because the Dance-the-Music requires music to be played at a strict tempo, it is easy to calculate where the (BPs) are situated. The description of the discrete dance steps itself is aimed towards the spatial deployment of gestures performed by the feet and body-center. The description contains two components: First, the spatial displacement of the body-center and feet, and second, the rotation of the body around the vertical axis.

Spatial displacement This parameter describes the time-dependent displacement (i.e., spatial segment) of the body-center and feet from one beat point (i.e., BP_{begin}) to the next one (i.e., BP_{end}) relative to the posture taken at the time of BP_{begin}. With posture, we indicate the position of the body-center and both feet at a discrete moment in time. Moreover, this displacement is expressed with respect to the local coordinate system (see Section 3.1.1) defined at BP_{begin}. In general, the algorithm executes the calculation in three steps:

1. *Input of absolute (x, y, z) values of body-center and feet at a sample rate of* 100 Hz.

2. *Calculation of the (x, y, z) displacement relative to the posture taken at BP_{begin} expressed in the global coordinate system (see Equation* 1*):*

   → For this, at the beginning of each step (i.e., at each BP_{begin}), we take the incoming absolute (x, y, z) value of the body-center and store it for the complete duration of the step. At each instance of the step-trajectory that follows, this value is subtracted from the absolute position values of the body-center, left foot, and right foot. This operation places the body-center at each BP_{begin} in the middle of the global coordinate system. As a consequence, this "reset" operation results in jumps in the temporal curves forming separate spatial segments corresponding each to one dance step (e.g., Figure 6, bottom). The displacement from the posture taken at each BP_{begin} is still expressed in an absolute way (i.e., without reference to the body). Therefore, the algorithm needs to perform a final operation.

3. *Rotation of the local coordinate system in a way it has the same orientation as the global coordinate system at BP_{begin} (cf., Section* 3.1.1*, step* 1*):*

   → Similar to the previous step, only the orientation of the rigid body attached to the body-center at each new BP_{begin} is taken into account and used successively to execute the rotation of all the following samples belonging to the segment of a particular step.

4. *Calibration:*

   → Before using the Dance-the-Music, a user is asked to take a default calibration pose, meaning to stand up straight with both feet next to each other. The (x, y, z) values of the feet obtained from this pose are stored and used to subtract from the respective coordinate values of each new incoming sample. As such, the displacement of the feet is described at each moment in time in reference to that pose. This calibration procedure enables to compensate for (1) individual differences in leg length, and (2) changes in the placement of the rigid bodies corresponding to the body-center. As such, one can opt to place that rigid body somewhere else on the torso (see Figure 2).

$$(\Delta x, \Delta y, \Delta z)_{[BP_i, BP_{i+1}[} = (x, y, z) - (x, y, z)_{BP_i} \tag{1}$$

Rotation According to Euler's rotation theorem, any 3-D displacement of a rigid body whereby one point of the rigid body remains fixed, can be expressed as a single rotation around a fixed axis crossing the fixed point of the rigid body. Such a rotation can be fully defined by specifying its quaternions. A quaternion representation of a rotation is written as a normalized four-dimensional vector $[q_x\ q_y\ q_z\ q_w]^T$, linked to the rotation axis $[e_x\ e_y\ e_z]^T$ and rotation angle ψ.

In Section 3.1.1, we outlined the reasons why the rotation of the rigid body attached to the body-center is restricted to rotations around the vertical axis without having too severe consequences for the freedom of dance performances. This is also an important aspect with respect to the calculation of the rotation around the vertical axis departing from quaternion values. Every rotation, expressed by its quaternion values, can then be approximated by a rotation around the vertical axis $[0\ 0\ \ \pm 1]^T$ or in aeronautics terms rotations are limited to *yaw*. Working with only yaw gives us the additional benefit of being able to split-up a dance movement in a chain of rotations where every rotation is specified with respect to the orientation at the beginning of each step (i.e., at each BP). The calculation procedure consists of two steps:

1. *Calculation of the rotation angle around the vertical axis:*

   → The element q_w in the quaternion (q_x, q_y, q_z, q_w) of the rigid body attached to the body-center determines the rotation angle ψ ($q_w = cos(\psi/2)$. We use this rotation angle as an approximation value for the rotation angle around the vertical axis (i.e., yaw angle Ψ). Implicitly, we suppose that the values for q_x and q_y are small meaning that the rotation axis approximates the vertical axis: $[e_x\ e_y\ e_z]^T = [0\ 0\ \pm 1]^T$.

2. *Calculation of the rotation angle relative to the orientation at BP_{begin} (see Equation 2):*

→ The method to do this is similar to the one described in the second step of the previous paragraph ('Spatial displacement').

$$\Delta\Psi_{[BP_i,BP_{i+1}[} = \Psi - \Psi_{BP_i} \tag{2}$$

3.2 Modeling of dance figures

In this section, we outline how we apply a template-based approach for modeling a sequence of repetitive dance figures performed on music. The parameters of the—what we will call—*basic step model* are the ones described in Section 3.1.2, namely the relative displacements of the body-center and feet, and the relative rotation of the body in the transverse plane per individual dance step.

The basic step model is considered as a spatiotemporal representation indicating the spatial deployment of gestures with respect to the temporal beat pattern in the music. The inference of the model is conceived as a supervised machine-learning task. In supervised learning, the training data consists of pairs of input objects and a desired output value. In our case, the training data consists of a set of p repetitive cycles of a specific dance figure of which we process the gestural parameters as explained in Section 3.1.2. The timing variable is the input variable and the gestural parameters are the desired values. The timing variable depends on (1) the number of steps per dance figure, (2) the tempo in which the steps are performed, and (3) the sample rate of the incoming raw movement data, according to Equation 3.

$$n = \frac{60 * \text{Steps per Figure*sample rate (Hz)}}{\text{tempo (bpm)}} \tag{3}$$

As such, the temporal structure of each cycle is defined by a fixed number of samples (i.e., 1 to n). The result is a single, fixed-size template of dimension $m{\times}n{\times}p$, with m equal to the number of gestural parameters (cf., Section 3.1.2), n equal to the number of samples defining one dance figure (cf., Equation 3), and p equal to the number of consecutive cycles performed of the same dance figure (see Figure 6).

To model each of the gestural parameters, we use a dedicated *K-Nearest Neighbor regression calculated with L1 loss function.* In all these models, *time* is the regressor. The choice for an L1 loss function ($L1 = |Y - f(t))|$) originates in its robustness (e.g., protection against data loss, outliers, etc.). In this case the solution is the conditional median, $f(t) = median(Y|T = t)$ and its estimates are more robust

compared to an L2 loss function solution that reverts to the conditional mean [30, p. 19–20]. We calculate the median of the displacement values and rotation value located in the neighborhood of the timestamp we want to predict for. Since we have a fixed number of sequences per timestamp (i.e., p) a logical choice is to choose all these values for nearest neighbor selection. The "K" - in the K-nearest neighbor selection is then determined by the number of sequences performed of the dance figure. The model that eventually will be stored as reference model consists of an array of values, one for each timestamp (see Figure 6).
Because the median filtering is applied sample per sample, it results in "noisy" temporal curves. Tests have proven that smoothing the temporal curves stored in the template improve the results of the recognition algorithms described in Section 3.3. Therefore, we smooth the temporal curves of the motion parameters of the model template with a Savitzky-Golay FIR filter (cf., [31]). This is done segment per segment to preserve the "reset" operation applied during the processing of the motion parameters (see Section 3.1.2). This type of smoothing has the advantage of preserving the spatial characteristics of the original data, like widths and heights, and it is also a stable solution.

The system is now able to model different dance figures performed on specific musical pieces and, subsequently, to store the basic step models in a database together with the corresponding music. In what follows, we will refer to these databases as *dance figure/music databases*. One singular database is characterized by dance figures which consist of an equal amount of dance steps performed at the same tempo. However, as many databases as one pleases can be created varying with respect to the amount of dance steps and tempi. These databases can then be stored as .txt files and loaded again afterwards. Once a database is created, it becomes possible to (1) visualize the basic step models contained in it, and (2) compare a new input dance performance with the stored models and provide direct audiovisual feedback on the quality of that performance. These features are described in the remaining part of this section on the technical design of the Dance-the-Music.

3.3 Dance figure recognition

The recognition functionalities of the Dance-the-Music are intended to estimate the quality of a student's performance in relation to a teacher's model. It is the explicit goal to help students to learn to imitate the teachers' basic step models as closely as possible. Therefore, the recognition algorithms are implemented to provide a measure of similarity (for individual motion features or for the overall performance). This measure is then used to give students feedback about the quality of their performance. For example, the dance-based music retrieval service presented in Section 3.4.2 must be conceived from this perspective.

In this section, we outline the mathematical method for estimating in real time the similarity between new movement input and basic step models stored in a dance figure/music database. For this, we will use a template matching method. This means that the gestural parameters calculated from the new movement input will be stored in a single, fixed-size *buffer template*, which can then be matched with the templates of the stored models (see Figure 7). A crucial requirement of such a method is that it must compensate for small deviations from the model in space as well as in time (cf., [32]). Spatial deviations do not necessarily need to be considered as errors. A small deviation in space (movement is performed slightly more to the left or right, higher or lower, forward or backward) should not be translated into an error. Similar, a performance slightly scaled with respect to the model (bigger or smaller) should also not be considered as an error. using normalized root mean square error (NRMSE) as a means to measure error is not appropriate as it does punish spatial translation and scaling errors. A better indicator for our application is the Pearson product-moment correlation coefficient r. It measures the size and direction of the linear relationship between our two variables (input and model). A perfect performance would result in a correlation coefficient that is equal to 1, while a total absence of similarity between input gesture and model would lead to a correlation coefficient of 0. Timing deviations are compensated by calculating the linear relationship between the gestural input and model as a function of a time-lag (cf., cross-correlation). If we apply a time-lag window of i samples in both directions, then we obtain a vector of $i{+}1$ r values. The maximum value is then chosen and outputted as correlation coefficient for this model together with the corresponding time-lag. As such, we obtain an objective measurement of whether a dance performance anticipates or is delayed with respect to the model.

The buffer consists of a single, fixed-size template of dimension $m_{\times}n$, with m equal to the number of gestural parameters (cf., Section 3.1.2), and n equal to the number of samples defining one dance figure (cf., Equation 3). When a new sample - containing a value for each processed gestural parameter - comes in, the system needs a temporal reference indicating where to store the sample in the template buffer on the *Time* axis. For this, dance figures are performed on metronome ticks following a pre-defined beat pattern and tempo. As such, it becomes possible to send a timestamp along with each incoming sample (i.e., a value between 1 and n).

Because the buffer needs to be filled first, an input can only be matched properly to the models stored in a dance figure/music database after the performance of the first complete dance figure. From then on, the system will compare the input buffer with all the models at the end of each singular dance step. This results for each model in m r values, with m corresponding to the number of different parameters defining the model. From these m values, the mean is calculated and internally stored. Once a comparison with all models is made, the highest r value is outputted together with the number of the corresponding model. An example of this mechanism is shown in Figure 8. The dance figure/dance database is here filled with nine basic step models. From these nine models, the model corresponding with the r values indicated with thicker line width, is the model that at all times most closely relates to the dance figure of which the data is stored in the input buffer template. As such, this would be the correlation coefficient that is outputted by the system together with the model number.

3.4 Audiovisual monitoring of the basic step models and real-time performances

As explicated in Section 2, multimodal monitoring of basic step models and real-time performances is an important component of the Dance-the-Music. In the following two sections, we explain in more detail respectively the visual and auditory monitoring features of the Dance-the-Music.

3.4.1 Visual monitoring

The contents of the basic step models can be visually displayed (see Figure 9) as a kind of dynamic and real-time dance notation system. What is displayed is (1) the spatial displacement of the body-center and feet, and (2) the rotation of the body around the vertical axis from $\mathrm{BP}_{\mathrm{begin}}$ to $\mathrm{BP}_{\mathrm{end}}$. The visualization is dynamic in the way it can be played back in synchronization with the music on which it was originally performed. It is also possible to adapt the speed of the visual playback (but then, without sound). The display visualizes each dance step of a basic step model in a separate window. Figure 9 shows the graphical notation of an eight-step basic samba figure as performed by the samba teacher of the evaluation experiment presented in Section 4. The window at the left visualizes the direct feedback that users get from their own movement when imitating the basic step model represented in the eight windows placed at the right. On top of the figure, one can see the main interface for controlling the display features. The main settings involve transport functions (play, stop, reset, etc.), tempo settings, and body part selection.

The intent is to visualize the displacement patterns (i.e., spatial segments) of each step on a two-dimensional plane which represents the ground floor on which the dance steps were performed (see Figure 9). In other words, the displacement patterns are displayed on the ground plane and viewed from a top-view perspective. Altering the size of the dots of which the trajectories exist, enable us to visualize the third, vertical dimension of the displacement patterns. The red dots and purple trajectories define the displacement patterns of the right foot, the green dots and yellow trajectories the ones of the left foot, and the black dots and trajectories the ones of the body-center. The vague-colored dots represent the configuration of the feet and body-center relative to each other at the beginning of the step ($\mathrm{BP}_{\mathrm{begin}}$, the sharp-colored dots the configuration and the end of the step ($\mathrm{BP}_{\mathrm{end}}$). As can be seen, as a result of the segmentation procedure presented in Section 3.1.2, the position of the body-center is reset at each new $\mathrm{BP}_{\mathrm{begin}}$. The triangle indicates the orientation of the body around the vertical axis. Moreover, the orientation of the windows (and all the data visualized in it) needs to be understood in reference to the local reference frame of the dancer (see Figure 5). Initially, the orientation and positioning of each window with respect to the local frame is as indicated by the XY coordinate system visualized in the left window. However, when dance novices are using the visual monitoring aid, they can make the orientation of the movement patterns of the basic step model displayed in each window dependable on their own rotation at the beginning of each new step. This means that the XY coordinate system (and, with that, all data visualizing the model) is rotated in such a way that it coincides with the local frame of the dance novice. As such, the basic step model is visualized at each instance from a first-person perspective. This way of displaying information presents an innovative way of giving real-time instructions about how to move the body and feet to perform a step properly. Now, this information can be transferred to the dancer in different ways:

1. The most basic option is to display the interface on a screen or to project it onto a big screen. When a dance figure involves a lot of turns around the vertical axis, it is difficult to follow the visualization and feedback on the screen. An alternative display method provides a solution to this problem. It concerns the projection of the displacement information directly on the ground. We used this last approach in the evaluation study presented in Section 4 (see Figure 2).

2. An alternative method projects the windows one by one, instead of all eight windows at once (see Figure 10). The position and rotation of the window is thereby totally dependent of the position and rotation of the user at the beginning of each new dance step ($\mathrm{BP}_{\mathrm{begin}}$). A new window is then projected onto the ground at each $\mathrm{BP}_{\mathrm{begin}}$, as such that the centroid of the window coincides with the position taken by the person at that moment. The rotation of the window is then defined as explained above

in this section. Because of the reset function (see Section 3.1.2) applied to the data - which visualizes the position of the body-center at each $\mathrm{BP}_{\mathrm{begin}}$ in the center of the window - the visualization gets completely aligned with the user. The goal for the dancer is then to stay aligned in time with the displacement patterns visualized on the ground. If one succeeds, it means that the dance step was properly performed. This method could not yet be evaluated in a full setup. However, the concept of it provides promising means to instruct dance figures.

3.4.2 Auditory monitoring

There have been designed ample computer technologies that facilitate automatic dance generation/synthesis from music annotation/analysis [33–36]. The opposite approach, namely generating music by automatic dance analysis, is explored in the domain of gesture-based human-computer interaction [37–39] and music information retrieval [10]. We will follow this latter approach by integrating a dance-based music querying and retrieval component in the Dance-the-Music. However, it is important to mention that this component is incorporated not for the sake of music retrieval as such, but rather to provide an auditory feedback supporting dance instruction. Particularly, the quality of the auditory feedback gives the students an idea in real time how well their performance matches the corresponding teacher's model. As will be explained further, the quality of the auditory feedback is related to two questions: (1) Is the correct music retrieved corresponding to the dance figure one performs? (2) What is the balance between the music itself and the metronome supporting the timing of the performance?

After a dance figure/music database has been created (or an existing one imported) as explained in Section 3.2, a dancer can retrieve a stored musical piece by executing repetitive sequences of the dance figure that correlate with the basic step model stored in the database together with the musical piece. The computational method to do this is outlined in Section 3.3.

The procedure to follow in order to retrieve a specific musical piece is as follows. The input buffer template is filled from the moment the metronome - indicating the predefined beat-pattern and tempo - is activated. Because the system needs the performance of one complete dance figure to fill the input buffer template (see Section 3.3), the template matching operation is executed only from the moment the last sample of the first cycle of the dance figure arrives. The number of the model which is then indicated by the system as being the most similar to the input triggers the corresponding music in the database. To allow a short period of adaptation, the “moment of decision” can be delayed untill the end of the second or third cycle. The retrieval of the correct music matching a dance figure is only the first step of the auditory feedback. Afterwards, while the dancer keeps on performing the particular dance figure, the quality of the performance is scored by the system. The score is delivered by the correlation coefficient r outputted by the system. On the one hand, the score is displayed visually by a moving slider that goes up and down along with the r values. On the other hand, the score is also monitored in an auditory way. Namely, according to the score, the balance between the volume of the metronome and the music is altered. When $r = 0$, only the metronome is heard. In contrast, when the $r = 1$, only the music is heard without the support of the metronome. The game-like, challenging character is meant to motivate dance novices to perform the dance figures as good as possible.

A small test was conducted to evaluate the technical design goals of this feature of the Dance-the-Music in an ecologically valid context. Moreover, it functioned as an overall pilot test for the evaluation experiment presented in Section 4.

For the test, we invited a professional dancer (female, 15 years of formal dance experience) to our lab where the OptiTrack motion capture system was installed. She was asked to perform four different dance figures in a different genre (tango, jazz, salsa and hip-hop) on four corresponding types of music played at a strict tempo of 120 beats per minute (bpm). The figures consisted of eight steps performed at a tempo of 60 steps per minute. The dancer was asked to perform each dance figure five times consecutively. From this training data, four models were trained as explained in Section 3.2 and stored in a database together with the corresponding music. Afterwards, the dancer was asked to retrieve each of the four pieces of music one by one as explained above in this section. She performed each dance figure six times consecutively. Because the dancer herself provided the models, it was assumed that her performances of the dance figures during the retrieval phase would be quite alike. The data outputted by the template matching algorithm (i.e., the model that most closely resembles the input and the corresponding r value) was recorded and can be seen in Figure 11. We only took into account the last five performed dance figures as the first one was needed to fill the input buffer. The analysis of the data shows that the model that was intended to be retrieved was indeed always recognized as the model most closely resembling the input. The average of the corresponding correlation values r over all performances was 0.59 (SD = 0.18). This value is totally dependent on the quality of the performance of the dancer during the retrieval (i.e., recognition) phase in relation to her performance during the modeling phase. Afterwards we noticed that smoothing the data contained in the model and the data of the real-time input optimizes the detected rate of similarity. As such, a Savitzky-Golay smoothing filter (see Section 3.3) was integrated and used in the evaluation experiment presented in the following section. Nonetheless, the results of this test show that the technical aspects of the auditory monitoring part perform to the design goals in an ecologically valid context.

4 Evaluation of the educational purpose

In this section, we describe the setup and results of a user study conducted to evaluate if the Dance-the-Music system can help dance novices in learning the basics of specific dance steps. The central hypothesis is that students are able to learn the basics of dance steps guided by the visual monitoring aid provided by the Dance-the-Music application (see Section 2.2). A positive outcome of this experiment would provide support to implement the application in an educational context. A demonstration video containing fragments of the conducted experiment can be advised in a supplementary file attached to this article.

4.1 Participants

For the user study, three dance teachers and eight dance novices were invited to participate. The three teachers were all female with an average age of 27.7 years (SD = 1.5). One was skilled in jazz (11 years formal dance experience, 3 years teaching experience), another in salsa (15 years formal dance experience, 5 years teaching experience) and the last in samba dance (9 years formal dance experience of which 4 years of samba dance). The samba teacher had no real teaching experience but, due to her many years of formal dance education, was found competent by the authors to function as a teacher. The group of students consisted of four males and four females with an average age of 24.1 years (SD = 6.2). They declared not to have had any previous experience with the dance figures they had to perform during the test.

4.2 Stimuli

The stimuli used in the experiment were nine basic step models produced by the three dance teachers (see Section 4.3). Each teacher performed three dance figures on a piece of music corresponding to their dance genre (jazz, salsa, and samba). They were able to make their own choice of what dance figure to perform within certain limits. We asked the teacher to choose dance figures consisting of eight individual steps and to perform them at a rate of 60 steps per minute (the music had a strict tempo of 120 bpm). The nine basic step models can be viewed in a supplementary file attached to this article. They involve combinations of (1) displacement patterns of the feet relative to the body-center, (2) displacement patterns of the body in absolute space, and (3) rotation of the body around the vertical axis.

4.3 Experimental procedure

The experimental procedure is subdivided into three phases, following the three basic procedures of the demonstration-performance method (see Section 2).

Demonstration phase In the first phase, basic step models were inferred from the performances of the teachers. The three teachers were invited to come one by one to the lab where the motion capture system was installed. First, the concept of the Dance-the-Music was briefly explained to them. Then, they were equipped with IR-reflecting markers to enable us to use the motion capture system. After that, they were allowed to rehearse the dance figures on the music we provided. When they said to be ready, they were asked to perform each dance figure five times consecutively. From these five cycles of training data, a basic step model was inferred. Each dance teacher was asked to perform three dance figures, resulting in a total of nine basic step models.

Learning phase What follows is a learning phase during which students are instructed how to perform the basic step models provided by the teachers, only aided by the visual monitoring system (see Section 3.4.1). As in the previous phase, the students were invited one by one to the experimental lab. Also, they were informed about the concept of the Dance-the-Music and the possibilities of the interface to control the visual monitoring aid, which was projected onto the floor (see Figure 2). After this short introduction, they were equipped with IR-reflecting markers. Then, the individual students were given 15 min to learn a randomly assigned basic step model. During this 15 min learning phase, they could decide themselves how to use the interface (body part selection, tempo selection, automated rotation adaptation, etc.).

Evaluation phase In the last phase, it is evaluated how well the students' performances match the teachers' models. All eight students were asked to perform the studied dance figure five times consecutively. From these five cycles, the first is not considered in the evaluation to allow adaptation. The performance is done without the assistance of the visual monitoring aid. Movements were captured and pre-processed as explained in Section 3.1. The template matching algorithm (see Section 3.3) was used to obtain a quantitative measure of the similarity (i.e., correlation coefficient r) between the students' performances and the teachers' models. Because an r value is outputted at each $\mathrm{BP}_{\mathrm{begin}}$, we obtain in total 32 r values. The mean of these 32 values was calculated together with the standard deviation to obtain an average score

r for each student. Moreover, their performances were recorded on video in order that the teachers could evaluate afterwards the performed dance figures in a qualitative way. Also, after the experiment, students were asked to complete a short survey questioning their user experience. The questions concerned whether the students experienced pleasure during the use of the visual monitoring aid and whether they found the monitoring aid helpful to improve their dance skills.

4.4 Results

The main results of the user study are displayed in Table 1. Concerning the average measure of similarity (r) between the students' performances and the teachers' models, we observe a value of 0.69 (SD = 0.18). From a qualitative point of view, the average score given by the teachers to the students' performances in relation to their own performances is 0.79 (SD = 0.10). Concerning the students' responses to the question whether they experienced pleasure during the learning process, we observe an average value of 4.13 (SD = 0.64) on a five-point Likert scale. The average score indicating the students' opinion about the question whether the learning method helps to improve their dance skills resulted in an average value of 4.25 (SD = 0.46).

4.5 Discussion

For the interpretation of the results, it is difficult to generalize results in terms of statistically significance because of the relatively small number of participants (N = 8). Therefore, a more qualitative interpretation of the data seems more suitable. Although the sample number is relatively small, the average r of 0.69 (SD = 0.18) suggests a considerable improvement of dance skills among all subjects due to the visual monitoring aid facilitated by the Dance-the-Music. Moreover, the average of the standard deviation of r (M = 0.06, SD = 0.02) indicates that the individual performances of the students were quite consistent over time. These results are supported by the results of the scores teachers' gave—based on video-observation—to the students' performances (M = 0.79, SD = 0.10). What is also of interest is the observation of a linear relationship ($r = 0.50$) between the scores provided by the template matching algorithm of the Dance-the-Music and the scores provided by the teacher. Concerning the user experience, results suggest that students in general experience pleasure using the visual monitoring aid (M = 4.13, SD = 0.64). This is an important finding as the experience of pleasure can stimulate students to practice with the Dance-the-Music. Even more important is the finding that the students in general have the impression that the Dance-the-Music is capable of helping them to learn the basics of dance gestures (M = 4.25, SD = 0.46). This suggests that the Dance-the-Music can be an effective aid in music education.

5 General discussion

The results provided in Section 4 suggest that the Dance-the-Music is effective in helping dance students to learn basic step models provided by a dance teacher. Despite these promising results, some remarks need to be made. First, the sample number (N = 8) was relatively small. This implies that, for the moment, the results indicate only preliminary tendencies and can not be generalized yet. Second, although the basic step models involved combinations of displacement patterns of the feet and body and rotation around the vertical axis, the models were anyhow relatively easy. This was necessary because (1) the students had no earlier experience with the dance genre, and (2) because it was the first time they actually interacted with the visual monitoring aid (and we expect a learning curve for students to use and get used to the dynamic visual notation system). Therefore, in the future, it would be of interest to conduct a longitudinal experiment investigating whether it becomes possible to learn more complex dance patterns when one becomes more familiar with the notation system presented by the visual monitoring aid. Third, in line with the previous remark, a comment made by the teachers was that dancing involves more than displacement patterns of the feet and body and rotation of the body around the vertical axis. This is indeed a justified remark. However, as stated before (e.g., Section 1), the Dance-the-Music can easily import other motion features and integrate them into the modeling logic based on spatiotemporal motion templates. For example, in the evaluation experiment, we integrated the horizontal, (x, y) displacement of the body in absolute space as a supplementary parameter in the model and visualization aid. Apart from that, it must be stressed that the explicit intent of the Dance-the-Music—as it is presented in this article—is to provide a platform which can help students to learn the basics of dance gestures which can then be further refined by the dance teacher during dance classes. Because the visually monitoring aid can in principle also be used without a motion capture system, it can be useful for students to use the Dance-the-Music to rehearse certain dance figures and small sequences of dance figures at home before coming to dance class. As such, the time that teacher and students are together can be optimally spent without "losing" time teaching the basics to the students.

Technological realizations and innovations incorporated in the Dance-the-Music were developed explicitly from a user-centered perspective, meaning that we took into account aspects of human perception and action learning. For example, the visualization strategy is based on findings on the role of (1) segmentation of complex events [15] , and (2) a first-person perspective [16] for human perception and action learning. However, it must be added that a third-person perspective (e.g., a student watching the teacher performing) has its own benefits with respect to action learning [16]. Therefore, both perspectives must be considered as being complementary to each other. Moreover, the introduction of a novel method for modeling and recognition based on spatiotemporal motion templates, in contrast to techniques based on HMM, facilitate to take into account time-space dependencies that are of crucial importance in dance performances. We also took into account research findings stressing the importance of real-time feedback of one's performance [19,21–24]. Therefore, we developed a recognition algorithm—based on template matching techniques—that enables us to provide real-time, multimodal feedback of a student's performance in relation to a teacher's model. Another property that was essential in the design of the Dance-the-Music was the dynamic and user-configurable character. In essence, the Dance-the-Music is considered as a technological framework of which the content depends completely on the user (i.e., teacher and student). Users can propose their own dance figures, music, tempo, etc. Moreover, the Dance-the-Music facilitates to incorporate a broad spectrum of movements (absolute displacement, rotation, etc.). These two features distinguish the Dance-the-Music from most dance games available on the commercial market that provide mostly a fixed, built-in vocabulary of dance moves and music and also provide only a small action space. Because of its dynamic character, the Dance-the-Music can also have its benefits for motor rehabilitation purposes.

6 Conclusion

In this article, we presented a computational platform, called Dance-the-Music, that can be used in dance education to learn dance novices the basics of dance figures. The design of the application is considered explicitly from a user-centered perspective, meaning that we took into account aspects of human perception and action learning. Aspects that are of crucial importance involve (1) time-space dependencies in dance performances, (2) the importance of segmentation processes and a first-person perspective for action learning, (3) the effectiveness of direct, multimodal feedback, and (4) the design of a dynamic framework of which the content is completely dependent on the users' needs and wishes. Technologies have been presented to bring these conceptual approaches into practice. Moreover, an evaluation study suggested that the Dance-the-Music is effective in learning the basics of dance figures to dance novices.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

This study was carried out in context of the EmcoMetecca project funded by the Flemish government. The authors want to thank Raven van Noorden for her contribution to the three-dimensional visualizations, Ivan Schepers for his technical assistance and Annelies Raes, Orphee Sergyssels and Leen De Bruyn for their willingness to participate in the evaluation study.

References

1. S Brown, M Martinez, L Parsons, The neural basis of human dance. Cerebral Cortex **16**(8), 1157–1167 (2006)

2. M Leman, *Embodied Music Cognition and Mediation Technology* (MIT Press, Cambridge, MA, USA, 2007)

3. M Leman, L Naveda, Basic gestures as spatiotemporal reference frames for repetitive dance/music patterns in Samba and Charleston. Music Percept. **28**, 71–91 (2010)

4. L Naveda, M Leman, The spatiotemporal representation of dance and music gestures using topological gesture analysis (TGA). Music Percept. **28**, 93–111 (2010)

5. RI Godøy, M Leman, *Musical Gestures: Sound, Movement, and Meaning* (Routledge, New York, NY, USA, 2010)

6. Kahol K, Tripathi P, Panchanathan S, Automated gesture segmentation from dance sequences, in *Proc. 6th IEEE International Conference on Automatic Face and Gesture Recognition (FG)* volume=not specified; pages=883–888; publisher=IEEE Computer Society; location=Seoul, South Korea (2004)

7. A Ruiz, B Vachon, Three learning systems in the reconnaissance of basic movements in contemporary dance, in *Proc. 5th IEEE World Automation Congress (WAC)*, vol. 13, pp. 189–194, IEEE Computer society, Orlando, FL, USA (2002)

8. F Chenevière, S Boukir, B Vachon, Compression and recognition of spatio-temporal sequences from contemporary ballet. Int. J. Pattern Recogn. Artif. Intell. **20**(5), 727–745 (2006)

9. K Kahol, K Tripathi, Panchanathan S, Documenting motion sequences with a personalized annotation system. IEEE Multimedia **13**, 37–45 (2006)

10. F Bévilacqua, B Zamborlin, A Sypniewski, N Schnell, F Guédy, N Rasamimanana, Continuous realtime gesture following and recognition, in *Gesture in Embodied Communication and Human-Computer Interaction*, pp. 73–84 vol. 5394, Springer Verlag, Berlin, Heidelberg, Germany (2010)

11. C Bishop, *Pattern recognition and machine learning.* (Springer Science+Business Media LLC, New York, USA, 2009)

12. A Bobick, J Davis, The representation and recognition of action using temporal templates. IEEE Trans. Pattern Anal. Mach. Intell. **23**(3), 257–267 (2001)

13. F Lv, R Nevatia, M Lee, 3D human action recognition using spatio-temporal motion templates. Comput. Vision Human-Comput. Interact. 120–130 (2005)

14. M Müller, T Röder, Motion templates for automatic classification and retrieval of motion capture data. in *Proc. ACM/Eurographics Symposium on Computer Animation (SCA)* pp 137- 146, Eurographics Association, Vienna, Austria (2006)

15. J Zacks, K Swallow, Event segmentation. Curr. Direct. Psychol. Sci. **16**(2), 80–84 (2007)

16. P Jackson, A Meltzoff, J Decety, Neural circuits involved in imitation and perspective-taking. Neuroimage **31**, 429–439 (2006)

17. D Davcev, V Trajkovic, S Kalajdziski, Celakoski S, Augmented reality environment for dance learning, in *Proc. IEEE International Conference on Information Technology, Research and Education (ITRE)*, (2003), pp. 189–193

18. Nakamura A, Tabata S, Ueda T, Kiyofuji S, Kuno Y, Dance training system with active vibro-devices and a mobile image display. In *Proc. IEEE International Conference on Intelligent Robots and Systems (IROS)*, 3075–3080, IEEE Computer Society, Alberta, Canada (2002)

19. J Chan, H Leung, J Tang, T Komura, A virtual reality dance training system using motion capture technology. IEEE Trans. Learn. Technol. **4**(2), 187–195 (2010)

20. L Deng, H Leung, N Gu, Y Yang, Real-time mocap dance recognition for an interactive dancing game. Comput. Animat. Virt. W. **22**, 229–237 (2011)

21. D Hoppe, M Sadakata, P Desain, Development of real-time visual feedback assistance in singing training: a review. J. Comput. Assist. Learn. **22**(4), 308–316 (2006)

22. E Gibbons, Feedback in the Dance Studio. J. Phys. Edu. Recreat. Dance **75**(7), 1–6 (2004)

23. J Menickelli, The Effectiveness of Videotape Feedback in Sport: Examining Cognitions in a Self-Controlled Learning Environment. *PhD thesis*, Western Carolina University (2004)

24. S Hanrahan, R Mathews, Success in Salsa: students' evaluation of the use of self-reflection when learning to dance, in *Proc. of the Conference of Tertiary Dance Council of Australia (TDCA)*, pp. 1–12, Melbourne, Australia (2005)

25. K Kahol, P Tripathi, S Panchanathan, T Rikakis, Gesture segmentation in complex motion sequences, in *Proc. IEEE International Conference on Image Processing (ICIP)*, vol. 2, pp. 105–108, IEEE Computer Society, Barcelona, Spain, (2003)

26. H Yang, A Park, S Lee, Gesture spotting and recognition for human–robot interaction. IEEE Trans. Robot. **23**(2), 256–270 (2007)

27. T Artieres, S Marukatat, P Gallinari, Online handwritten shape recognition using segmental hidden markov models. IEEE Trans. Pattern Anal. Mach. Intell. **29**(2), 205–217 (2007)

28. S Rajko, G Qian, T Ingalls, J James, Real-time gesture recognition with minimal training requirements and on-line learning, in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, IEEE Computer Society, Minneapolis, USA (2007)

29. S Rajko, G Qian, HMM parameter reduction for practical gesture recognition, in *Proc. 8th IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, pp. 1– 6 IEEE Computer Society. Amsterdam, The Netherlands, (2008)

30. T Hastie, R Tibshirani, J Friedman, J Franklin, The elements of statistical learning: data mining, inference and prediction. Math. Intelligencer **27**(2), 83–85 (2005)

31. PJ Maes, M Leman, M Lesaffre, A model-based sonification system for directional movement behavior, in *Proc. 3th Interactive Sonification Workshop (ISon)*, (KTH, Stockholm, Sweden, 2010), pp. 91–94

32. F Lv, R Nevatia, Recognition and segmentation of 3-d human action using hmm and multi-class adaboost, in *Proc. 9th European Conference on Computer Vision (ECCV)*, vol. 3954, pp. 359–372 , Springer Verlag, Graz Austria, (2006)

33. H Mori, S Ohta, J Hoshino, Automatic dance generation from music annotation, in *Proc. International Conference on Advances in Computer Entertainment Technology (ACE)* ,pp. 352– 353 ACM Singapore, (2004)

34. T Shiratori, A Nakazawa, K Ikeuchi, Dancing-to-Music Character Animation. Comput. Graph. Forum **25**(3), 449–458 (2006)

35. J Kim, H Fouad, J Sibert, J Hahn, Perceptually motivated automatic dance motion generation for music. Comput. Animat. Virt. W. **20**(2–3), 375–384 (2009)

36. F Ofli, E Erzin, Y Yemez, A Tekalp, Multi-modal analysis of dance performances for music-driven choreography synthesis, in *Proc. IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*,pp. 2466–2469 IEEE Computer Society, Dallas, TX, USA, (2010)

37. G Qian, F Guo, T Ingalls, L Olson, J James, T Rikakis, A gesture-driven multimodal interactive dance system, in *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, vol. 3, pp. 1579–1582, IEEE Computer Society, Taipei, Taiwan (2004)

38. G Castellano, R Bresin, A Camurri, G Volpe, User-centered control of audio and visual expressive feedback by full- body movements. Affect Comput. Intell. Interact vol. 4738, pp 501- 510 (2007)

39. PJ Maes, M Leman, K Kochman, M Lesaffre, M Demey, The “One-Person Choir”: a multidisciplinary approach to the development of an embodied human-computer interface. Comput. Music J. **35**(2), 1–15 (2011)

Figure 1: **Graphical user interface (GUI) of the Dance-the-Music**.

Figure 2. **A student interacting with the interface of the visual monitoring aid, projected on the ground**.

Figure 3: **Schematic overview of the technical design of the Dance-the-Music**.

Figure 4: **Placement of the rigid bodies on the dancer's body**.

Figure 5. **Representation of how the body-fixed local coordinate system is translated to coincide with the global coordinate system**.

Figure 6. **Top left:** $m_{\times}n_{\times}p$ **template storing the training data**. Each cube consists of one numeric value which is a function of the time, gestural parameter and sequence. Top right: $m_{\times}n$ template representing a basic step model. Bottom left: The five lines represent an example of the contents of the gray cubes in the top left template (with $n = 800$, and $p = 5$). Bottom right: Representation of the discrete values stored in the gray feature array in the top right template.

Figure 7: **Template matching schematic**.

Figure 8. **Example of the internal mechanism of the template matching algorithm**. It represents the result of the comparison of a dance figure consisting of eight steps (defined each by 100 samples) performed by a student (here, subject 8 of the user study presented in Section 4) against all stored models (N=9) at each BP_{begin}.

Figure 9: **Visualization of the visual monitoring aid interface**.

Figure 10. **An example of how to project the eight windows one by one to create a real-time dance notation system incorporating a first-person perspective**.

Figure 11. r **values when the model outputted correspondingly is similar to the intended model**.

Table 1. Descriptive overview of the results of (1) the quantitative (A) and qualitative (B) ratings of similarity between students' performances and the corresponding teachers' models, and (2) the user experience of the dance students (C)

	Subjects	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Average**
	Age	25	24	28	24	20	33	12	27	
	Model	2	7	8	1	6	9	4	3	
A	Mean r	0.82	0.54	0.84	0.45	0.43	0.75	0.84	0.83	0.69 (SD = 0.18)
	SD r	0.03	0.03	0.04	0.07	0.10	0.05	0.06	0.07	0.06 (SD = 0.02)
B	Teacher's rating 0 = min, 1 = max	0.9	0.8	0.9	0.6	0.8	0.85	0.8	0.7	0.79 (SD = 0.10)
C	Pleasure	4	4	4	4	4	3	5	5	4.13 (SD = 0.64)
	Educational potential 5-point Likert scale 1 = strongly disagree, 5 = strongly agree	4	4	4	5	5	4	4	4	4.25 (SD = 0.46)

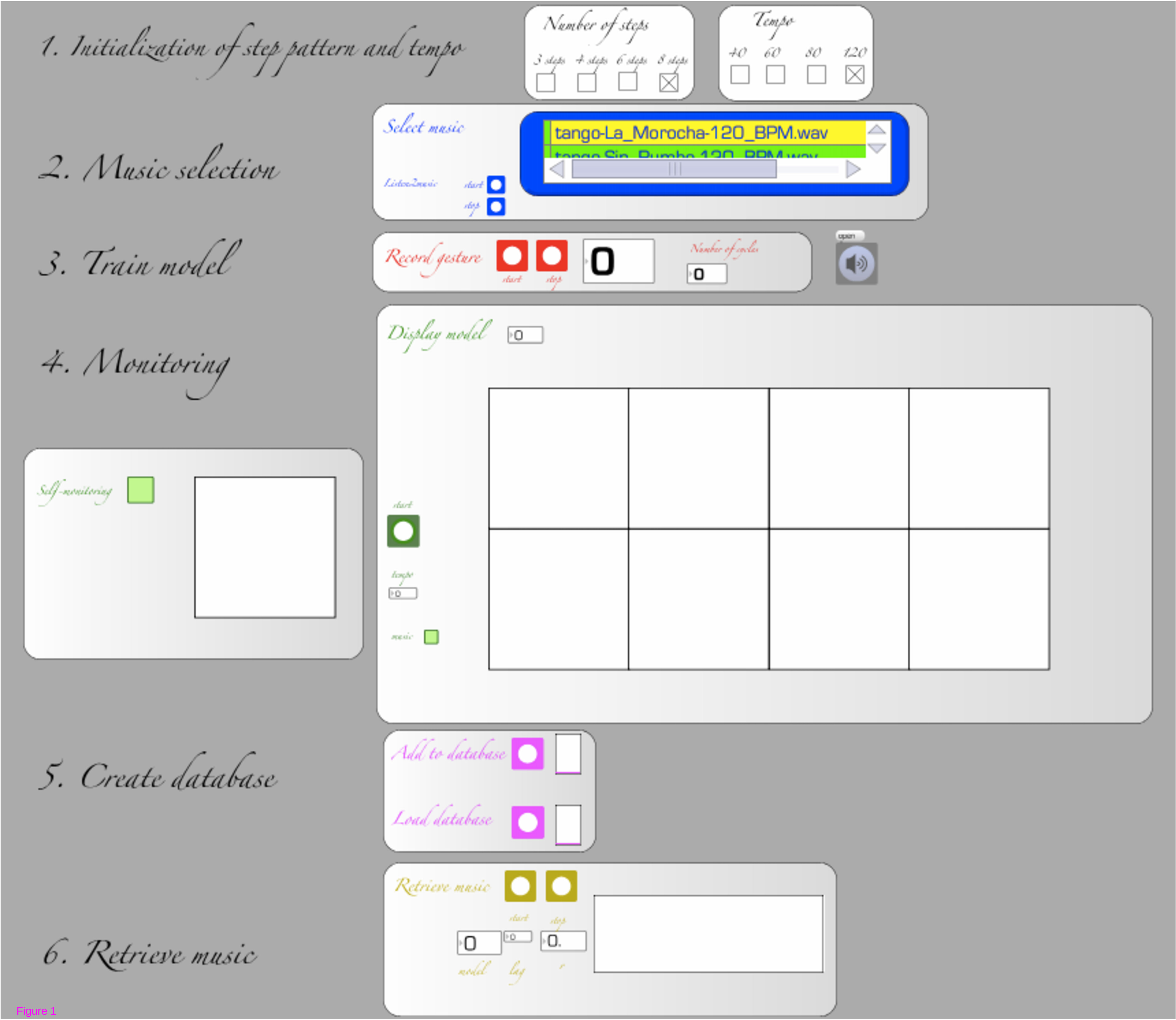
Additional Files

Additional file 1—AddFile1.pdf

Visualization of the nine basic step models proposed by the three dance teachers participating in the evaluation experiment presented in Section 4.

Additional file 2—DtMMaesPJ.mov

A demonstration video containing fragments of the evaluation experiment presented in Section 4.



1. Initialization of step pattern and tempo

Number of steps: 3 steps □ 4 steps □ 6 steps □ 8 steps ☒

Tempo: 40 □ 60 □ 80 □ 120 ☒

2. Music selection

Select music: tango-La_Morocha-120_BPM.wav
tango-Sin_Rumbo-120_BPM.wav

3. Train model

Record gesture 0

Number of cycles 0

4. Monitoring

Display model 0

tempo 0

5. Create database

6. Retrieve music

0 model
0 lag
0. r

Figure 1

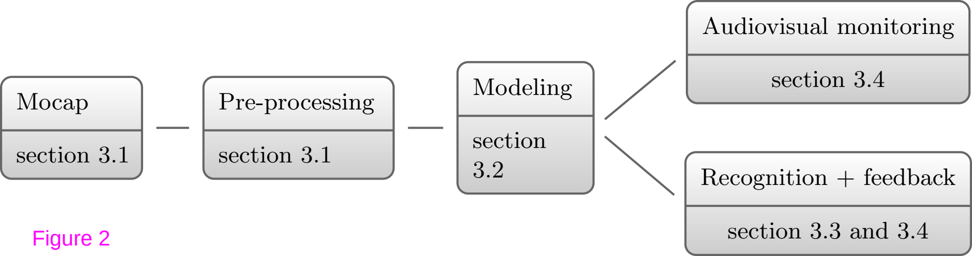
Mocap
section 3.1
Pre-processing
section 3.1
Modeling
section 3.2
Audiovisual monitoring
section 3.4
Recognition + feedback
section 3.3 and 3.4

Figure 2

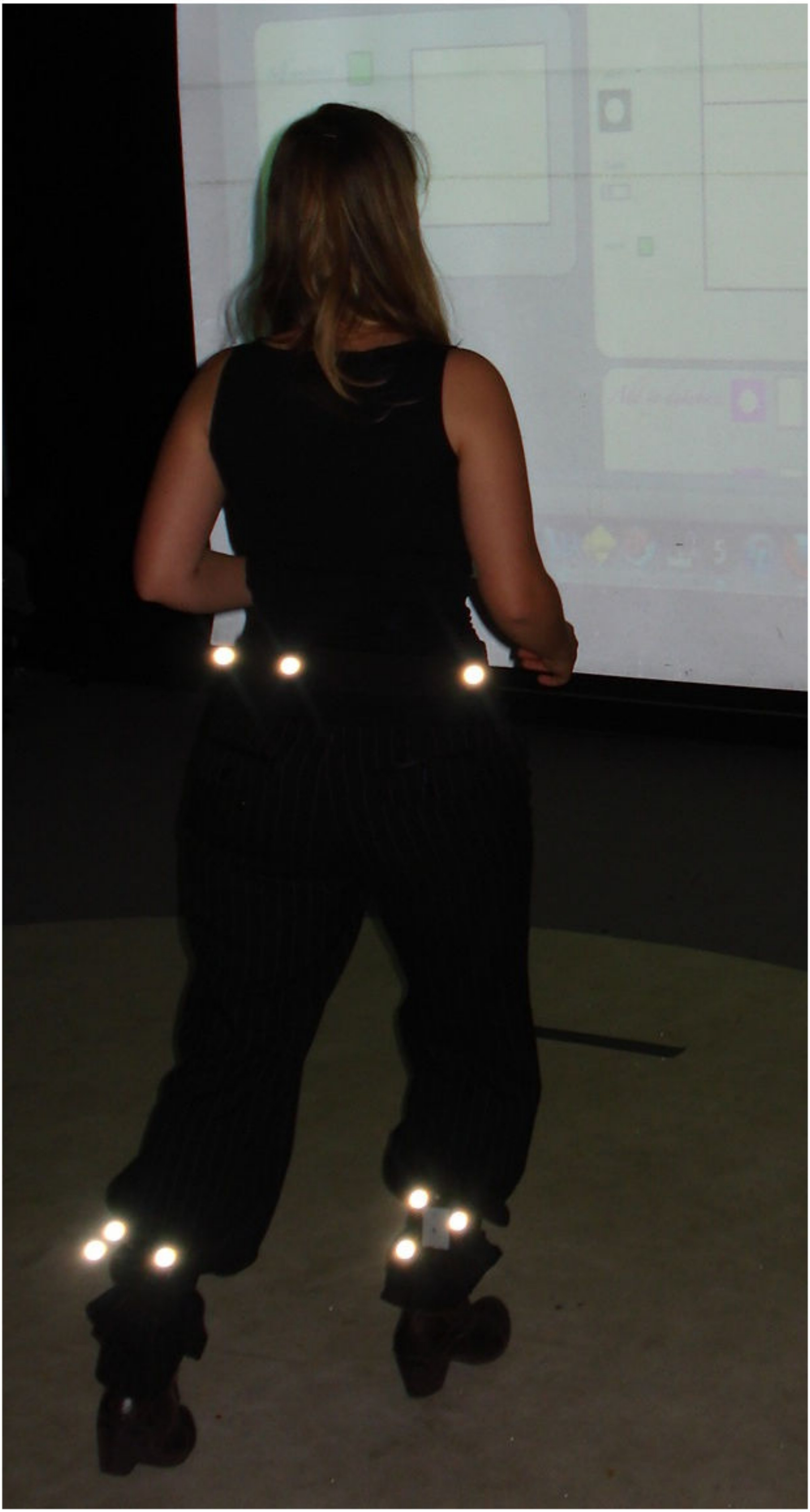
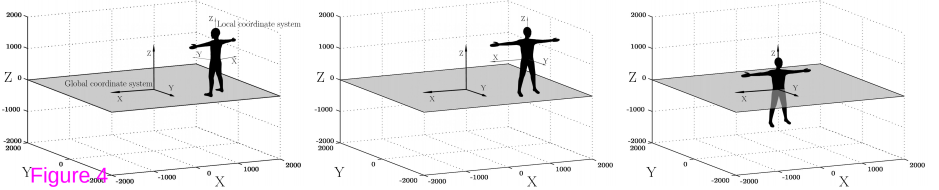


Figure 3

Figure 4


2000
1000
Z 0
-1000
-2000
2000
0
Y
-2000
-2000
-1000
0
X
1000
2000
Z
Local coordinate system
Y
X
Z
Global coordinate system
Y
X
2000
1000
Z 0
-1000
-2000
2000
0
Y
-2000
-2000
-1000
0
X
1000
2000
Z
X
Y
Z
Y
X
2000
1000
Z 0
-1000
-2000
2000
0
Y
-2000
-2000
-1000
0
X
1000
2000
Z
Y
X

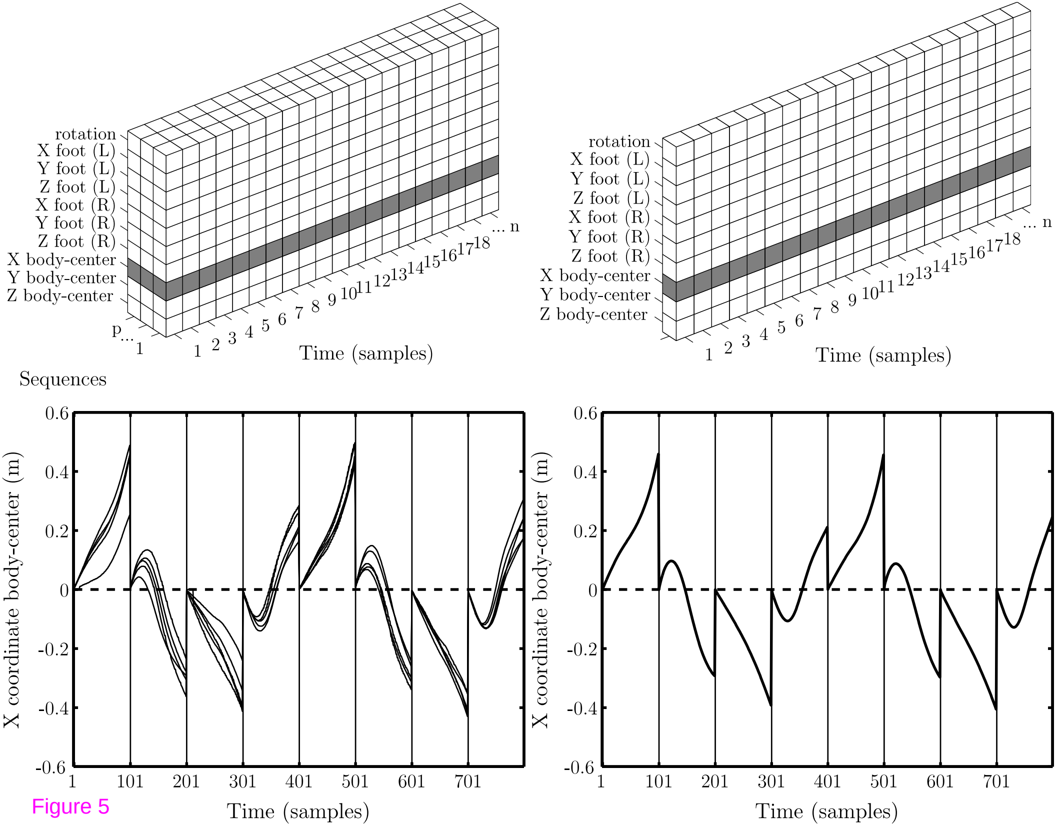


Figure 5

rotation
X foot (L)
Y foot (L)
Z foot (L)
X foot (R)
Y foot (R)
Z foot (R)
X body-center
Y body-center
Z body-center
p
...
1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 ... n
Time (samples)
Sequences
rotation
X foot (L)
Y foot (L)
Z foot (L)
X foot (R)
Y foot (R)
Z foot (R)
X body-center
Y body-center
Z body-center
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 ... n
Time (samples)
X coordinate body-center (m)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
1
101
201
301
401
501
601
701
Time (samples)
X coordinate body-center (m)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
1
101
201
301
401
501
601
701
Time (samples)

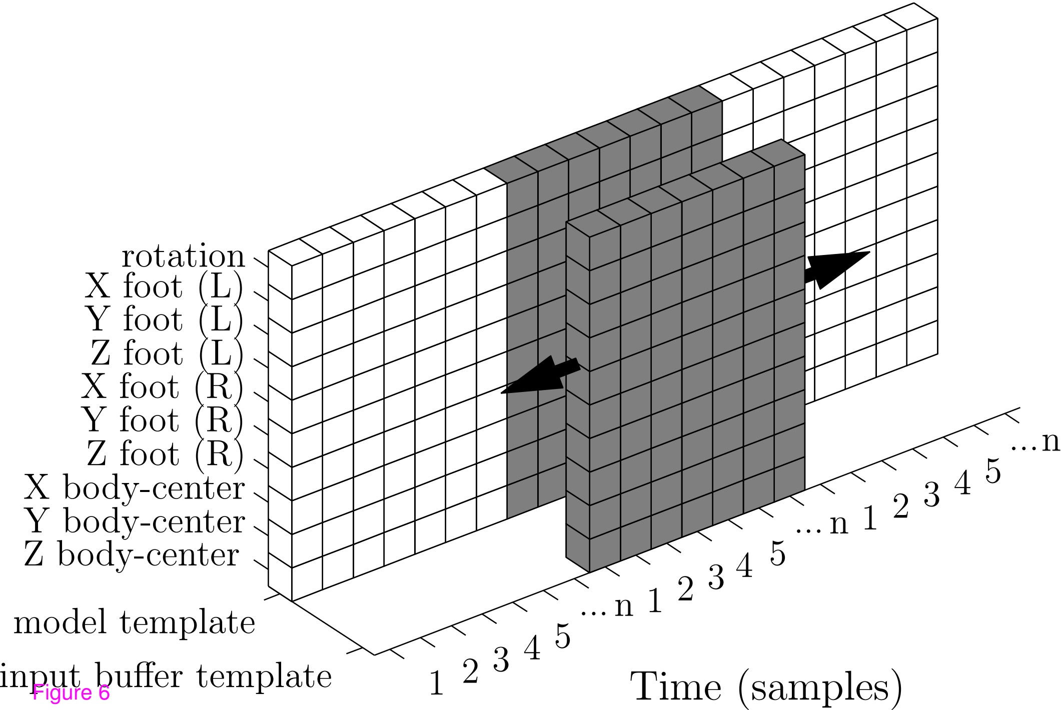

rotation
X foot (L)
Y foot (L)
Z foot (L)
X foot (R)
Y foot (R)
Z foot (R)
X body-center
Y body-center
Z body-center
model template
input buffer template
1 2 3 4 5 ... n 1 2 3 4 5 ... n 1 2 3 4 5 ... n
Time (samples)

Figure 6

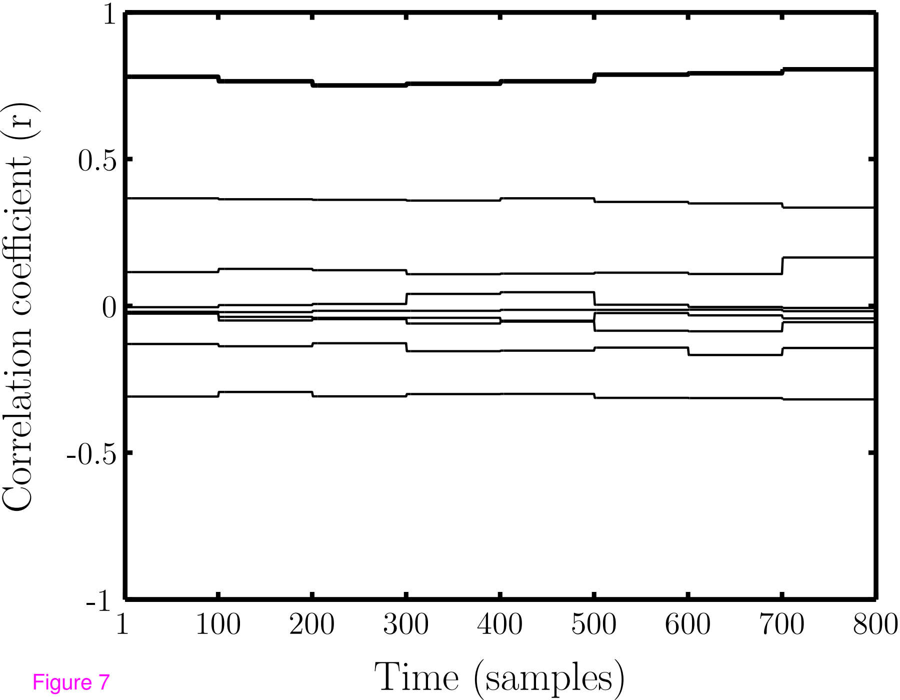
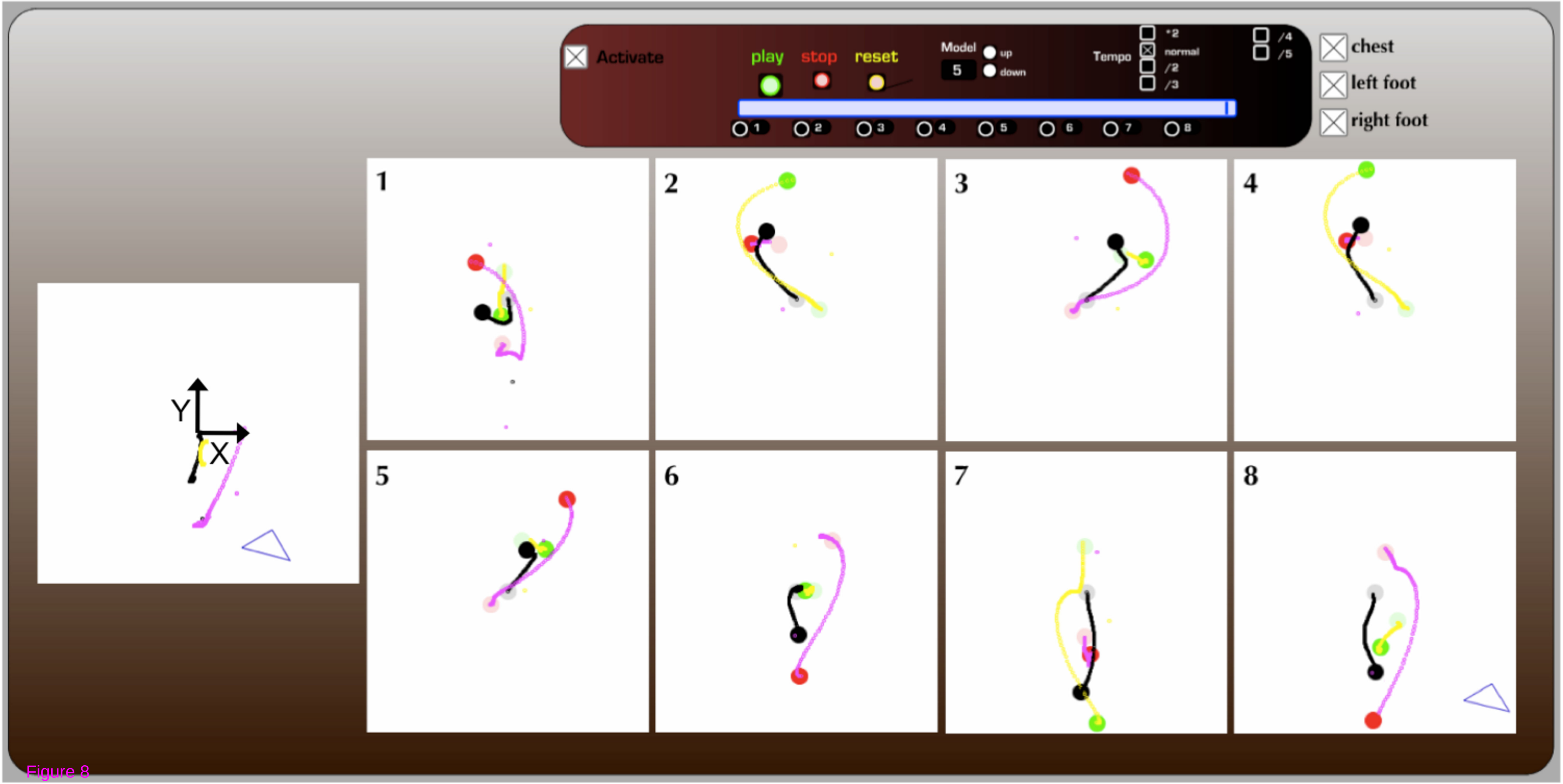

1
0.5
0
-0.5
-1
Correlation coefficient (r)
1
100
200
300
400
500
600
700
800
Time (samples)

Figure 7



☒ Activate

play ● stop ● reset ●

Model 5 ○ up ○ down

Tempo □ *2 ☒ normal □ /2 □ /3 □ /4 □ /5

☒ chest

☒ left foot

☒ right foot

Figure 8

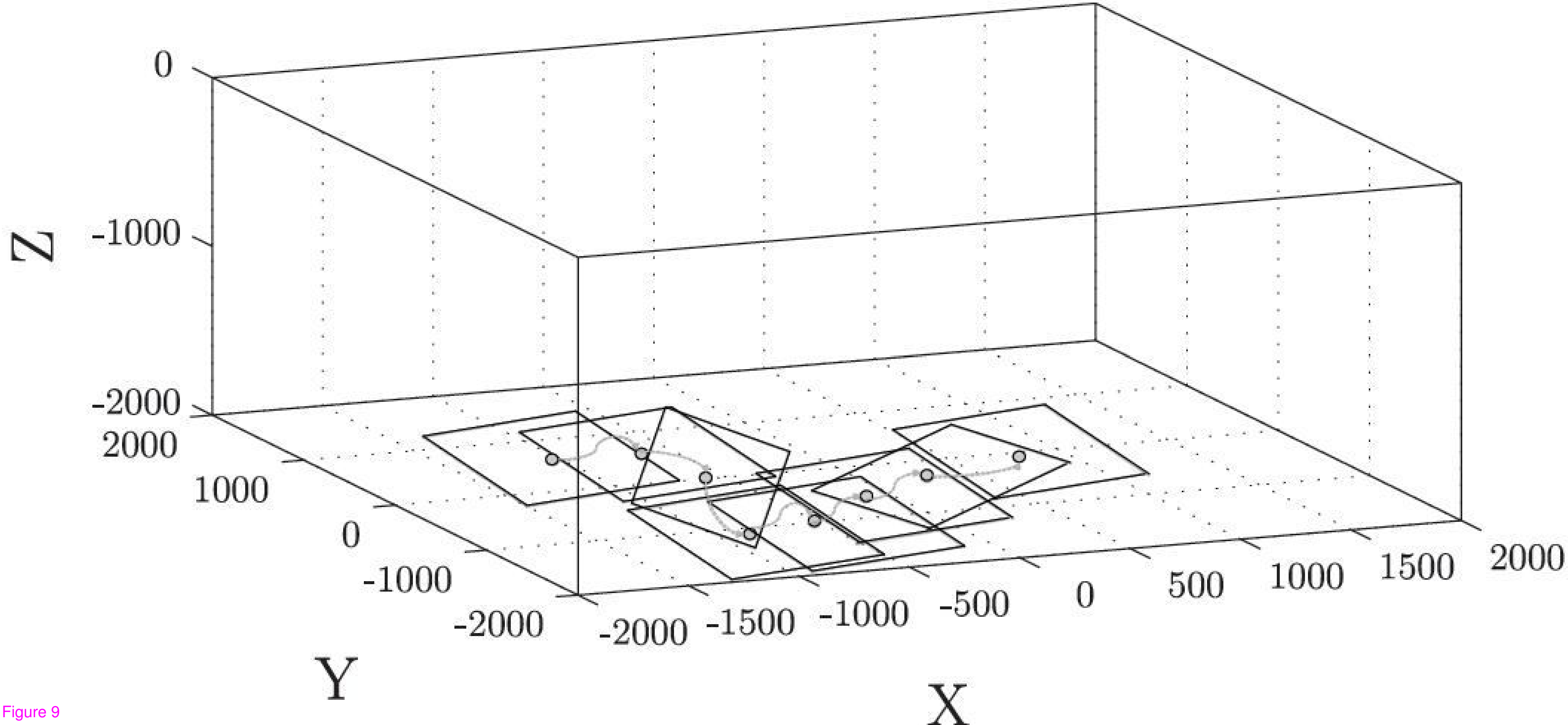

0
-1000
-2000
Z
2000
1000
0
-1000
-2000
Y
-2000
-1500
-1000
-500
0
500
1000
1500
2000
X

Figure 9

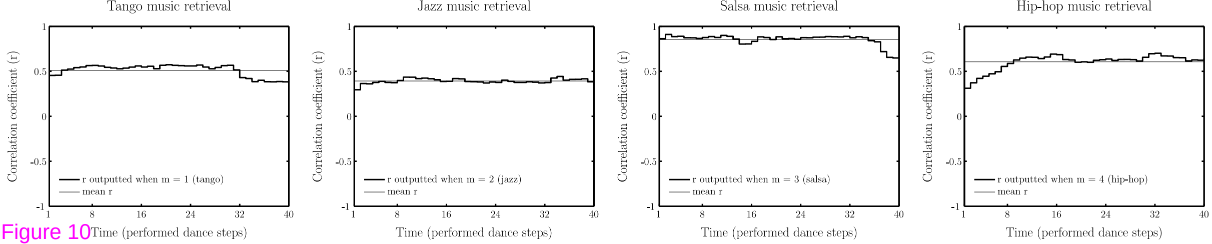

Tango music retrieval
Jazz music retrieval
Salsa music retrieval
Hip-hop music retrieval
Correlation coefficient (r)
1
0.5
0
-0.5
-1
1
8
16
24
32
40
Time (performed dance steps)
r outputted when m = 1 (tango)
r outputted when m = 2 (jazz)
r outputted when m = 3 (salsa)
r outputted when m = 4 (hip-hop)
mean r

Figure 10

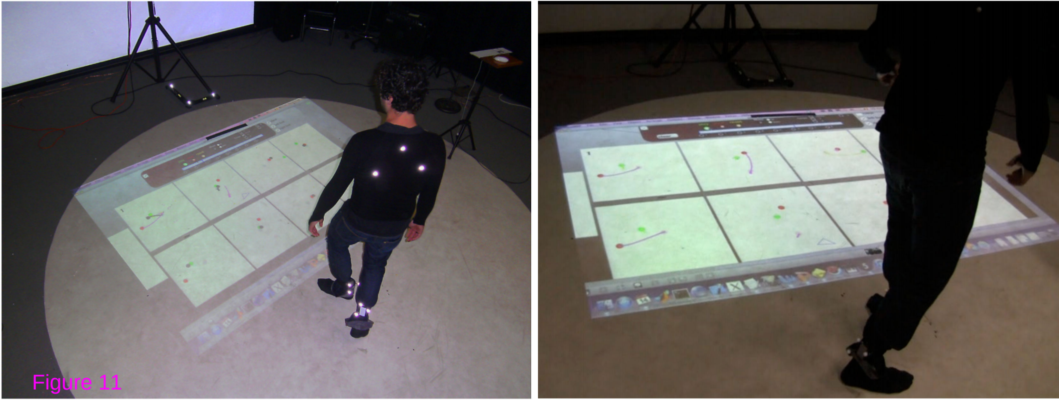


Figure 11

Additional files provided with this submission:

Additional file 1: AddFile1.pdf, 290K
http://asp.eurasipjournals.com/imedia/2754757696252975/supp1.pdf
Additional file 2: DtMMaesPJ.divx, 18353K
http://asp.eurasipjournals.com/imedia/2769525076252976/supp2.divx