Perceptually motivated automatic dance motion generation for music

By Jae Woo Kim*, Hesham Fouad, John L. Sibert and James K. Hahn
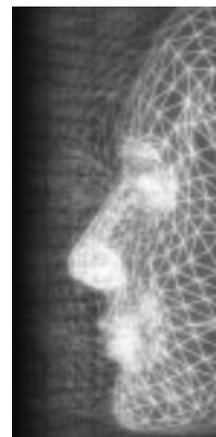
In this paper, we describe a novel method to automatically generate synchronized dance motion that is perceptually matched to a given musical piece. The proposed method extracts 30 musical features from musical data as well as 37 motion features from motion data. A matching process is then performed between the two feature spaces considering the correspondence of the relative changes in both feature spaces and the correlations between musical and motion features. Similarity matrices are introduced to match the amount of relative changes in both feature spaces and correlation coefficients are used to establish the correlations between musical features and motion features by measuring the strength of correlation between each pair of the musical and motion features. By doing this, the progressions of musical and dance motion patterns and perceptual changes between two consecutive musical and motion segments are matched. To demonstrate the effectiveness of the proposed approach, we designed and carried out a user opinion study to assess the perceived quality of the proposed approach. The statistical analysis of the user study results showed that the proposed approach generated results that were significantly better than those produced using a random walk through the dance motion database. Copyright © 2009 John Wiley & Sons, Ltd.

Introduction

Synthesizing realistic human motion is one of the most important research topics in computer animation. Various methods such as keyframing, inverse kinematics, and dynamic simulation have been used to synthesize human behavior. More recently, with the advent of motion capture technology, motion capture and editing techniques have become widely used in realistic human motion synthesis.

On the other hand, synchronization of sound with motion is also a very essential problem in animation because sound plays an important role in computer animation. The sound design process is tedious, time consuming, and requires a high level expertise to produce convincing results. Automatic sound effects synchronization has therefore been a crucial issue within the computer animation research community as well as industry.

When dealing with human dance motion, the synchronization between music and motion becomes even more important because dance motion has a much stronger linkage to music than any other type of motion. However, synchronizing music to dance motion is a very difficult problem due to the intricate relationship that exists between music and motion in a dance performance. To date, little research has been done on the problem of synchronization between dance motion and music.[1]

The synchronization problem is further compounded due to the complexity inherent in both musical and human motion data. Musical sound contains a wealth of information such as pitch, timbre, rhythm and, harmony. Human motion data, on the other hand, are multidimensional and of all human motion (i.e., walking, jumping, running); human dance motion is the most complex. This complexity makes it difficult to analyze the data and to explore the relationship between the musical and dance motion features.

---

*Correspondence to: J. W. Kim, Computer Science Department, The George Washington University, 801 22nd Street NW, Washington, DC 20052, USA. E-mail: zoltar@gwu.edu

The nature how dance is created and performed also creates additional challenges. Dance performance is carefully choreographed by expert choreographers based on a given musical piece. This process requires a high degree of intelligence and requires much expertise, education, and experience. This problem is therefore not amenable to the use of analytic or algorithmic models because of the aesthetic, perceptual, and psychological aspects that are involved in this process.

In this paper, we develop a new method to automatically generate synchronized dance motion that is perceptually matched to a given musical piece so that the resulting dance performance is convincing. The approach suggested in this paper can be applied to a number of application areas including film, TV commercials, virtual reality applications, computer games, and entertainment systems.

Related Work

There have been some recent efforts addressing the problem of synchronization between music and motion. Those efforts are focused on establishing a correlation between musical and motion features that represent the perceptual properties of music and motion, respectively.

Motion and music can be synchronized by matching events extracted from musical data with events extracted from motion data. An event in the musical domain is defined as a point in the time domain where a certain significant perceptual change occurs. Events in the musical domain include: dominant drum beats, peak points in amplitude while events in the motion domain include: motion beats, footsteps, arm swings, sudden pauses, and jumps.[2–5] Sauer and Yang suggested an approach to match musical events such as beat positions and dynamics (e.g., peaks and valleys of amplitude) with pre-defined actions.[5] Kim et al.[3] proposed an approach to synchronize motion to music by an incremental time-warping process that aligns the motion beats and the musical beats. Alankus et al.[2] suggested an approach for synthesizing dance motion by recombining dance moves to match musical beats. Lee and Lee suggested an approach to generate background music from an animation by matching feature points extracted from musical data and the corresponding feature points extracted from motion data.[4]

Another approach that has been used for music-to-motion synchronization is based on the features of music and motion. Musical features and motion features are parameters that describe the perceptual properties of music and motion. Events, on the other hand, represent the occurrences of some changes of features in time domain. Dobrian and Bevilacqua suggested a method of transforming motion features into either MIDI parameters or control parameters for signal processing based on a user specified mapping rules.[6] Shiratori et al.[7,8] suggested a feature based method for synthesizing synchronized dance motion based on rhythm and intensity of motion and music.

Research efforts in music and motion synchronization that consider the emotional responses of audiences to music and human motion are also considered here.[1] Cardle et al. suggested an approach for imbuing generated human motion with emotional content based on input music by manually establishing a mapping between musical features and motion editing filters interactively. Morioka et al.[9] suggested an algorithm of synthesizing music that can appropriately express emotion in dance by matching music and motion in emotional state space.

The research efforts that we investigated in this section have several important limitations when considering the problem of automatic dance motion generation. First, a limited number of features extracted from music and human motion were used in the matching process. The lack of richness in the feature set used in those efforts limits the effectiveness of the matching processes because not enough information is used and thus the matching are generally not convincing. Second, global structures of music and motion were not considered in matching process. Music and dance motion sequences consisted of several patterns or themes which change and repeat over the performance. This global structure of music and dance motion is important and provides critically important information that must be considered if convincing dance motion is to be generated. Previous efforts, however, focused exclusively on local matching process, ignoring the global structure of the music and motion.

Dance Performance Generation

The problem addressed in this paper is the automatic generation of human dance performances based on an arbitrary musical input. As depicted in Figure 1, the proposed approach constructs a motion graph consisting of human dance motion capture data in a pre-processing phase. When the motion graph is constructed, motion feature vectors are calculated for each motion segment
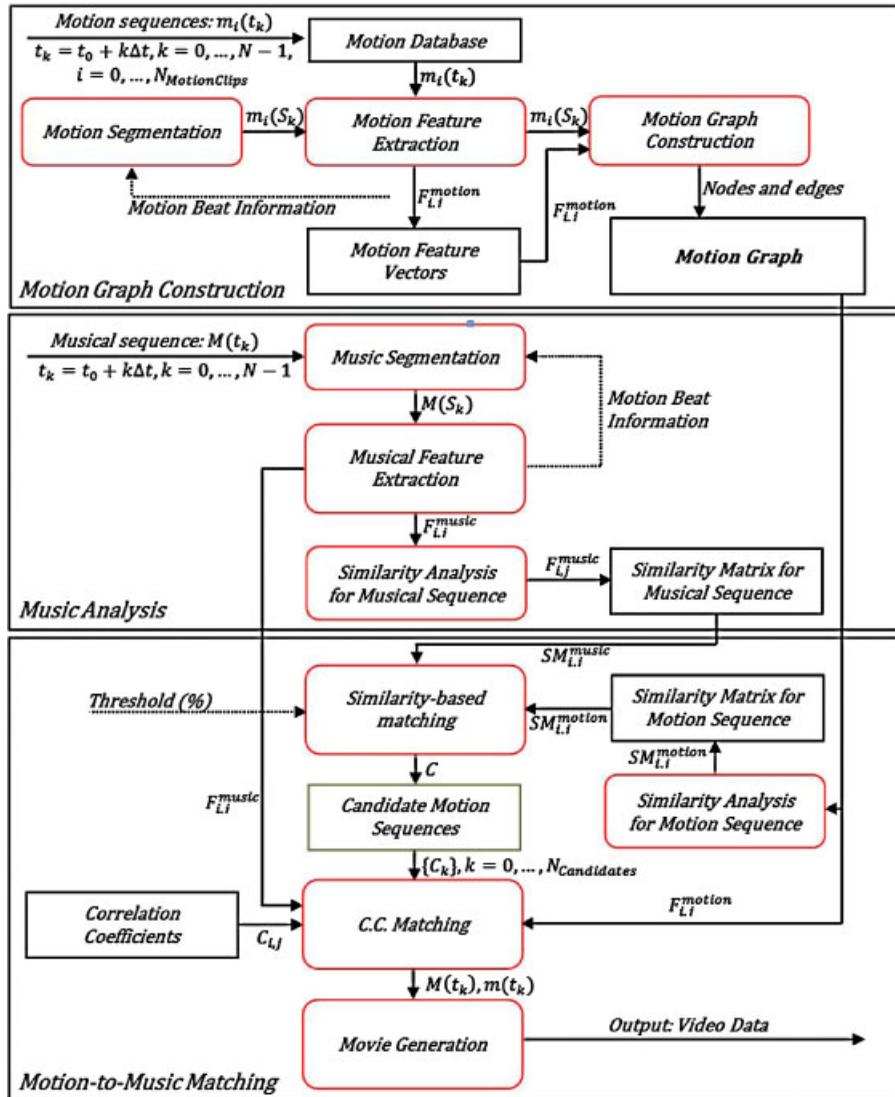
Figure 1. System overview for automatic dance motion generation.

and stored in the corresponding nodes of the graph. At run time, a musical piece is fed into the system and a music analysis process obtains a sequence of musical feature vectors by analyzing the input musical signal. Finally, a matching process between the musical feature vectors and motion feature vectors is performed by searching the motion graph to select an optimal path whose motion feature vector sequence best matches to the musical feature vector sequence of the input musical piece.

A motion graph used for dance motion synthesis in this work was constructed by using the method suggested by Reference [10]. Human dance motion can

therefore be synthesized by searching for a path within the motion graph that best satisfies a given criteria. The criteria for searching the path used in this research are based on the degree of matching between the sequence of motion feature vectors and the sequence of musical feature vectors. In this research, the motion graph nodes contain dance motion sequences consisting of 16 musical beats. We assume that 16 musical beats is an appropriate minimal length that can contain a dance motion.

The music analysis process extracts 30 musical features categorized into three parts: rhythm, pitch, and timber information. Music analysis is carried out on each musical segment. The size of each segment is

based on musical beat information where each segment consists of 16 beats. In this research, we used the musical features defined in Tzanetakis' work.[11]

Motion Analysis

The motion analysis process analyzes a motion segment and extracts a set of motion features which can describe the perceptual properties of the motion resulting in a motion feature vector. The motion feature vector contains useful information which can be used in matching the motion with musical features extracted in music analysis component.

Dynamic Features. Dynamic features represent the dynamic properties of the human motion. These features are conceptually related to some of the effort components of Laban Movement Analysis. Dynamic features used in this paper include: velocity, acceleration, and directional change of joint motion.[12]

Motion velocity is defined as a linear summation of the velocities of each joint position. Equation (1) depicts the calculation of motion velocity.

$$\text{motion\_velocity} = \sum_{j=1}^{N} \sum_{i=1}^{n} \|x_i(j+1) - x_i(j)\| \quad (1)$$

Here, $x_i(j)$ represents a position vector of a joint i at frame j and $x_i(j+1) - x_i(j)$ represents the change of position of joint i at frame j. n is the number of joints and N is the number of frames. Shiratori *et al.*[7,8] used a similar motion feature with this and they call it motion intensity. They found that motion intensity is perceptually related with music intensity. For example, fast motion matches well with strong musical sound and slow motion matches well with calm or heavy mood musical sound.

Like motion velocity, motion acceleration is defined as a linear sum of the accelerations of each joint position. Acceleration is proportional to and estimates the force applied to the joint. For example, dynamic or sudden motion will result in large acceleration values while static smooth motion results in low acceleration values. It is expected that high tempo, dynamic music will result in motion with high acceleration values. Motion acceleration is calculated by Equation (2).

$$\text{motion\_acceleration} = \sum_{j=1}^{N} \sum_{i=1}^{n} \|v_i(j+1) - v_i(j)\| \quad (2)$$

Here, $v_i(j)$ represents a velocity of joint i at frame j and obtained by calculating $\|x_i(j+1) - x_i(j)\|$.

Certain motions are simple or linear while others are complicated and consist of many directional changes. Directional change of motion indicates the degree of directional changes made by each joint of the human body. Consequently, motion with a high degree of directional change will have a large high frequency component. Directional change in motion is a significant feature in dance motion and would normally correlate to properties of the accompanying musical piece. The directional change of motion is calculated by Equation (3).

$$\text{directional\_change} = \sum_{j=1}^{N} \sum_{i=1}^{n} \cos^{-1}\left(\frac{x_i(j+1) \cdot x_i(j)}{\|x_i(j+1)\| \, \|x_i(j)\|} \right)$$

$$(3)$$

Here, $x_i(j)$ and $x_i(j+1)$ are vectors of two consecutive joint positions in time. Therefore, directional change of motion is obtained by calculating linear summation of angles two consecutive vectors make.

Postural Features. Postural features represent the properties that describe the shape of the human body movement. Postural features are conceptually similar to some of shape components of Laban Movement Analysis.[12] Postural features used in this paper are motion span, motion density, arm shape, footsteps, and balance. In general, the lower body of human character structure represents locomotion while the upper body represents gestures. Vertical, horizontal, and sagittal components of arm shape represent the characteristics of character's upper body gesture and footsteps provide information about lower body locomotion. Motion span, motion density, and balance give useful information about overall shape of motion.

Motion span represents the size of a motion based on the amount of space it spans. Motion span is calculated by considering how far each joint position travels in a motion segment. It is approximated by calculating the linear summation of the distance between each joint position in a trajectory of a joint relative to the calculated centroid of all the joint positions for that joint. Motion span is calculated as follows:

$$c_i = \frac{1}{N} \left(\sum_{j=1}^{N} x_i(j) \right) \quad (4)$$

$$motion\_span = \sum_{i=2}^{n} \sum_{j=1}^{N} \|x_i(j) - c_i\| \qquad (5)$$

where c_i is a vector representing the centroid of all the positions of joint i through the motion.

Motion density represents how dense the motion is. When a joint moves a great deal in a relatively small region, we define this as dense motion. Conversely, if it moves little in a relatively large region we define this as sparse motion. Motion density can therefore be calculated as the ratio of motion velocity over motion span as follows:

$$motion\_density = \frac{\sum_{i=2}^{n} \sum_{j=1}^{N} \|x_i(j+1) - x_i(j)\|}{\sum_{i=2}^{n} \sum_{j=1}^{N} \|x_i(j) - c_i\|} \qquad (6)$$

Arm shape features have three components according to the axes of the local Cartesian coordinate attached to the virtual human body. They are: vertical, horizontal, and sagittal components. The vertical component describes how much the arms are raised upward or lowered downward, the horizontal component describes the degree to which the arms are spread outward or contracted inward in the horizontal plane, and finally the sagittal component describes the degree to which the arms are extended forward or backward.[13]

Footsteps and balance provide critical information in describing dance motion. The footsteps feature is defined as the number of steps occurring in a given motion frame. Exuberant and flamboyant dance motions tend to have frequent steps while austere and restrained dance motions tend to have less frequent steps. Human body posture can be well balanced and stable or unbalanced and unstable. The postural feature balance is defined as the degree of stability of the human body posture and it provides useful information describing the posture of the dancer's movement. We calculate the projection point of the center of the human body onto the ground and the center point of the rectangle defined by the positions of the left and right heels. The feature balance is defined as the distance between those two positions.[9]

Nine dynamic features and 28 postural features are used to construct a motion feature vector by applying statistics such as mean, variance, and range on the motion features defined in this section. Those features are calculated for each dance motion segment and stored in corresponding graph nodes when the motion graph is constructed.

Motion-to-Music Matching

Automatic dance motion generation is, in effect, a problem of searching and selecting the optimal motion sequence from the motion graph that best matches an input musical piece. Motion-to-music matching deals with the problem of comparing two instances from different spaces—musical feature space and motion feature space. If there existed a standard space that is shared by both of the two instances, the comparison (or match) would be apparent: we can measure the distance between two points in a single standard space. We, however, do not have such a standard space and we therefore have to compare the two points in two different spaces.

The idea of the suggested approach is that although the two points under comparison are in different spaces, we can at least match the relative amount of perceptual changes in auditory and visual sensations. While musical and motion segments proceed from one segment to another during a performance, the audience can percieve the changes of auditory and visual stimuli. We assume that the difference in auditory sensation between two consecutive musical segments must be matched to the difference in the visual sensation between the corresponding two motion segments. For example, if there is a drastic change such as calm sound to noisy and strong sound, it is expected that the corresponding motion segments also change from, for example, slow and smooth motion to fast and dynamic motion. Otherwise, if there is not much change in auditory sensation, then not much change in visual sensation is expected.

We use the similarity matrices to match the amount of relative perceptual changes in auditory and visual sensations. The similarity matrix is a matrix which represents the similarities among segments in music or motion sequences. As shown in Figure 2, columns and rows of the similarity matrix represent the segments and a cell in the matrix represents the similarity between the corresponding segments. The similarity between the segments is obtained by calculating the normalized Euclidean distance between the two points of the segments in the feature space. Similarity matrices are calculated for both the music and motion sequences. Equation (7) shows the calculation of the similarity between music and motion sequences.

$$similarity_{i,j} = \sqrt{\sum_{k=0}^{n-1} (F_k^i - F_k^j)^2} \qquad (7)$$
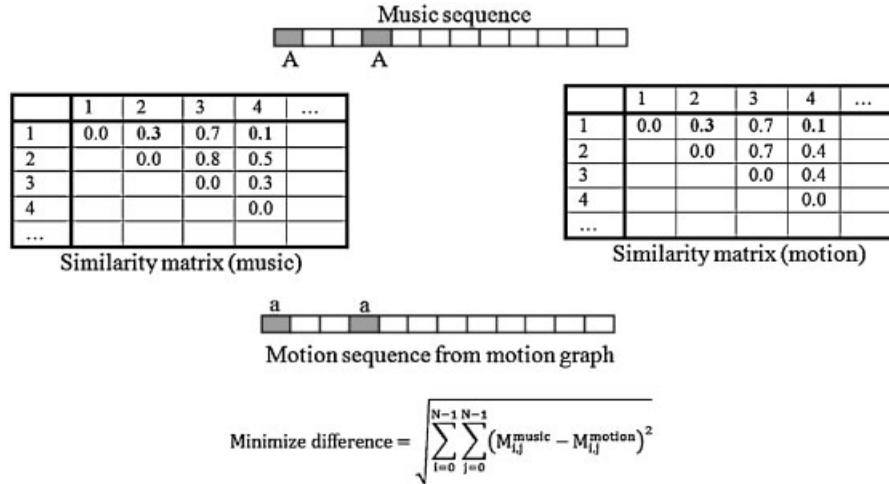
Figure 2. Matching progressions of musical and motion patterns using similarity matrices.

Here, F_i^k represents the normalized kth feature value of the feature vector for the segment i and n is the number of features. The same equation applies to both the music segments and the motion segments.

After obtaining the similarity matrices for both of the music and motion sequences, we calculate the difference between the two matrices to evaluate the quality of current matching. Equation (8) shows the calculation of the difference between the similarity matrices for music and motion sequences. The matching which minimizes the difference is determined to be the best match.

$$\text{difference} = \sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(M_{i,j}^{\text{music}} - M_{i,j}^{\text{motion}})^2} \qquad (8)$$

Both music and dance performances contain repeated patterns. We assume that when a musical pattern repeats, the dance motion pattern similar to the corresponding dance motion that was previously matched to that musical pattern should be repeated. This can be achieved by minimizing the difference between the similarity matrices of musical and motion sequences as described above so that the progressions of musical and motion patterns are matched.

Although the approach that minimizes the similarity matrices effectively matches the progressions of patterns in music and the generated dance sequences, it is, however, not sufficient to produce good results because it does not consider the correlation between the features of the music and dance sequences. In our approach, the matching of progressions of musical and motion patterns generates a set of candidate motion sequences.

We then establish the correlation between musical and motion features to finally select the best matching motion sequence.

We use correlation coefficients to measure the strength of correlation between each pair of musical and motion features. Equation (9) shows the calculation of correlation coefficient between ith musical feature and jth motion feature, $r_{i,j}$. Here, x_i denotes the ith musical feature and y_j denotes the jth motion feature. The calculation is carried on an example data set that consists of the pairs of musical and motion data that have been shown to have excellent perceptual matching in a subjective user study.

$$r_{i,j} = \frac{n\sum x_i y_j - (\sum x_i)(\sum y_j)}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_j^2 - (\sum y_j)^2}} \qquad (9)$$

Human perception is known to be highly sensitive to changes in visual and auditory sensations. We must therefore consider the correlation between the changes in perception of both musical and motion contents when performing the matching process. To do this, we calculate the correlation coefficients on changes in perception in corresponding pairs of music and motion segments. More specifically, we compute the correlation coefficients between Δx_i and Δy_j. Here, Δx_i denotes the difference of a ith musical feature between two consecutive musical segments and Δy_j denotes the difference of a jth motion feature between corresponding two motion segments.

$$r_{j,k}^{\Delta} = \frac{n\sum \Delta x_j \Delta y_k - (\sum \Delta x_j)(\sum \Delta y_k)}{\sqrt{n\sum \Delta x_j^2 - (\sum \Delta x_j)^2}\sqrt{n\sum \Delta y_k^2 - (\sum \Delta y_k)^2}}$$

$$(10)$$

In the matching process, for every musical-motion feature pairs that have strong correlation, the difference between the motion feature value and expected motion feature value that is determined by a regression function is calculated. For example, if ith musical feature x_i and jth motion feature y_j has a strong correlation, then the expected motion feature $F_{k,j}^{'\text{motion}}$ is obtained by Equation (11). Here, $F_{k,j}^{\text{motion}}$ denotes jth motion feature of motion segment k and the function $R(\cdot)$ is a regression function obtained from regression analysis on the example data set.

$$F_{k,j}^{'\text{motion}} = R(F_{k,i}^{\text{music}}) \qquad (11)$$

The difference $D_{k,(i,j)}$ between the real motion feature value $F_{k,j}^{\text{motion}}$ and the expected motion feature $F_{k,j}^{'\text{motion}}$ can be obtained by the Equation (12).

$$D_{k,(i,j)} = \|F_{k,j}^{'\text{motion}} - F_{k,j}^{\text{motion}}\| = \|R(F_{k,i}^{\text{music}}) - F_{k,j}^{\text{motion}}\| \qquad (12)$$

The difference values $D_{k,(i,j)}$ are summed for every (i, j) pair that have strong correlation to assess the degree of matching between a musical segment and its corresponding motion segment. This calculation is carried out for all the musical and motion segment pairs of the musical and motion sequences under comparison and summed to evaluate the quality of matching between the sequences. Finally the motion sequence that minimizes the summed difference value is selected as the best match. The same process is carried out on all the pairs of Δx_i and Δy_j that have strong correlation.

User Feedback

It is difficult to evaluate dance performances using objective mathematical or algorithmic approaches. We therefore performed a subjective user opinion study to validate the effectiveness of the approach suggested in this paper. The study presented here tests whether the suggested approach can generate more creative, realistic dance performances that are better matched to the accompanying music than a random. Random here implies a process of randomly selecting a sequence of motion segments and aligning the beats of the dance sequence with the input music.

This study will test the following predictions:

- *Prediction* 1: The suggested approach generates a more creative dance performance than the random approach.

- *Prediction* 2: The suggested approach generates a more realistic dance performance than the random approach.
- *Prediction* 3: The suggested approach generates a dance performance that expresses the characteristics of the input music better than the random approach.
- *Prediction* 4: The overall impression of the results generated by the suggested approach is better than that of the results generated by the random approach.

Methodology

The participants of the study consisted of 20 people, 14 were male and 6 were female. Ages range from 15 to 42 and the average age was 26. All the participants were volunteers and they were not paid. Eight dance performances were shown to each participant. The study consisted of two separate sessions: one was a training session and the other was an evaluation session. Two of the eight dance performances were used in the training session and the other six were used in the evaluation session.

There were two categories of dance performance. The first represents dance performances generated by the suggested approach while the second represents dance performances generated randomly. Each category consists of three dance performances generated with input music with a different tempo—slow, intermediate, and fast. The stimuli therefore consist of six conditions in total. The six dance performances used in the evaluation session were randomly ordered to minimize any order effect.

During the study, participants were asked to view six video clips containing dance performances with accompanying music. Each video clip takes about 25 seconds. After viewing each video clip, they were asked to answer four questions related with the four hypotheses described above by scoring the quality of the dance performance they saw using seven-level Likert scale. We compared the results using analysis of variance to test for significance and the analysis results are shown in Table 1.

Discussion

The null hypotheses were rejected at the 0.05 or 0.01 level in all but 2 of the 12 cases. In all cases the score was higher for our approach, implying that there is a difference in response to the dance clips and that the approach presented here produces better results than a random one.

computer animation
& virtual worlds

Prediction 1: Dance performance generated by suggested approach looks more creative. The difference was significant for fast ($F = 17.273$, $p = 0.0 < 0.001$) and slow ($F = 17.300$, $p = 0.0 < 0.001$) music. The null hypothesis, that the results were the same, was rejected.

Musical tempo	Suggested approach	Random walk	*F*-value	*p*-value
Fast	5.5	3.5	17.273	<0.001
Intermediate	4.65	4.6	0.015	0.902
Slow	5.55	3.85	17.300	<0.001

Prediction 2: Dance performance generated by suggested approach looks more realistic. The difference was significant for fast ($F = 4.560$, $p = 0.039 < 0.05$), intermediate ($F = 6.818$, $p = 0.013 < 0.05$), and slow ($F = 6.818$, $p = 0.013 < 0.05$) music. The null hypothesis, that the results were the same, was rejected.

Musical tempo	Suggested approach	Random walk	*F*-value	*p*-value
Fast	6.2	5.6	4.560	0.039
Intermediate	6.35	5.6	6.818	0.013
Slow	6.4	5.85	6.818	0.013

Prediction 3: Dance performance generated by suggested approach looks more expressive. The difference was significant for fast ($F = 76.382$, $p = 0.000 < 0.001$), intermediate ($F = 5.240$, $p = 0.028 < 0.05$), and slow ($F = 115.024$, $p = 0.000 < 0.001$) music. The null hypothesis, that the results were the smae, was rejected.

Musical tempo	Suggested approach	Random walk	*F*-value	*p*-value
Fast	6.45	4.45	76.382	<0.001
Intermediate	5.75	4.7	5.240	0.028
Slow	6.05	3.1	115.024	<0.001

Prediction 4: Dance performance generated by suggested approach looks better overall, The difference was significant for fast ($F = 13.977$, $p = 0.001 < 0.05$) and slow ($F = 42.309$, $p = 0.000 < 0.001$) music. The null hypothesis, that the results were the same, was rejected.

Musical tempo	Suggested approach	Random walk	*F*-value	*p*-value
Fast	6.1	4.75	13.977	0.001
Intermediate	5.5	4.95	3.386	0.074
Slow	6.15	4.35	42.309	<0.001

Table 1. Statistical analysis of user opinion study results

For us the most important prediction is that the suggested approach can generate a dance performance that is perceptually well matched (expresses the music) to the input music. We also observed that people tend to think that the dance performances are more realistic and creative when they are perceptually well matched to the music. This result was somewhat unexpected because our approach does not have any functionality to generate "creative" dance movements and all the dance motions are realistic because they had been captured

from real dancer's movements. It may be that there is an underlying positive response (possibly aesthetic) that causes this correlation in the results.

We also noted that the two cases where the null hypothesis could not be rejected both correspond to the intermediate tempo and that the ratings for the intermediate tempo are closest in all four cases. This could be an interaction effect with tempo or it could have something to do with the participants overall experience with dancing.

Overall we feel encouraged by the study and think it suggests that our approach is promising.

Conclusions

In this paper, we proposed a novel approach for the generation of dance performances based on a given musical piece by matching the progressions of musical and motion patterns and by correlating musical and motion features. To do this, we introduced similarity matrices for musical and motion sequences and matched the progressions of musical and motion contents by minimizing the difference between the two similarity matrices. We used correlation coefficients to measure the strength of correlation between each pair of the musical and motion features and correlations between musical and motion features were established by matching the musical and motion feature pair that showed strong correlations. By doing this, the progressions of musical and dance motion patterns and perceptual changes between two consecutive musical and motion segments were matched. The statistical analysis of a user opinion study showed that the proposed approach is feasible and can generate natural, realistic, pleasing, and perceptually appropriate dance motion for a given piece of music.

References

1. Cardle M, Barthe L, Brooks S, Robinson P. Music-driven motion editing: local motion transformations guided by music analysis. In *Proceedings of Eurographics UK Chapter—EGUK Conference. Eurographics*, 2002; 38–44.
2. Alankus G, Bayazit AA, Bayazit OB. Automated motion synthesis for dancing characters: motion capture and retrieval. *Computer Animation and Virtual Worlds* 2005; **16**(3–4): 259–271.
3. Kim TH, Park SI, Shin SY. Rhythmic-motion synthesis based on motion-beat analysis. *ACM Transactions on Graphics* 2003; **22**(3): 392–401.
4. Lee H-C, Lee I-K. Automatic synchronization of background music and motion in computer animation. *Computer Graphics Forum* 2005; **24**(3): 353–362.
5. Sauer D, Yang YH. Music-driven character animation. *ACM Transactions on Multimedia Computing, Communications and Applications*. http://www.cs.ualberta.ca/~yang/animation_files/Sauer_Yang.pdf.
6. Dobrian C, Bevilacqua F. Gestural control of music: using the vicon 8 motion capture system. In *NIME'03: Proceedings of the 2003 Conference on New Interfaces for Musical Expression*, 2003; 161–163.
7. Shiratori T, Nakazawa A, Ikeuchi K. Dancing-to-music character animation. *Computer Graphics Forum* 2006; **25**(3): 449–458.
8. Shiratori T, Nakazawa A, Ikeuchi K. Synthesizing dance performance using musical and motion features. In *ICRA*, IEEE, 2006; 3654–3659.
9. Morioka H, Nakatani M, Nishida S. Proposal of an algorithm to synthesize music suitable for dance. In *ACE '04: Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, ACM, New York, NY, USA, 2004; 296–301.
10. Kovar L, Gleicher M, Pighin FH. Motion graphs. In *SIGGRAPH '02: Proceedings of the 29th annual Conference on Computer Graphics and Interactive Techniques*, ACM, New York, NY, USA, 2002; 473–482.
11. Tzanetakis G. Manipulation, Analysis and retrieval systems for audio signals. *PhD Thesis*, Princeton, NJ, USA, 2002 (Adviser-Perry Cook).
12. Laban R. *The mastery of movement*, 3rd edn revised by Lisa Ullmann. Play Inc, Boston, 1980.
13. Chi D, Costa M, Zhao L, Badler N. The emote model for effort and shape. In *SIGGRAPH'00: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 2000; 173–182.

Authors' biographies:

Jae Woo Kim is a doctoral student of Computer Science, The George Washington University and he has been working for VRSonic Inc. since 2003. Previously, he worked for Electronics and Telecommunications Research Institute (ETRI) in Korea where he worked in the area of neural networks, image processing, and auditory information processing. He received his M.S. in computer science in 1993 and B.S. in physics in 1991 from Hankuk University of Foreign Studies, Korea. He is currently working on music visualization, dance performance generation and affective state detection and induction for training simulation.

Dr Hesham Fouad is a veteran of the software industry. He began his career at IBM in an advanced technology center working to develop a multimedia subsystem for

the OS/2 operating system. His interest in audio led him to pursue doctoral degree in Computer Science with an emphasis on real-time, parallel systems for media generation. His dissertation research focused on perceptually optimized, real-time synthesis of virtual sonic environments. After completing his degree, Dr Fouad founded VRSonic in order to address the need for next generation technologies and tools for the design and synthesis of complex virtual sonic environments. Today, VRSonic works with government, academia, and industry to provide spatial audio solutions for training, entertainment, and research.

John L. Sibert is Professor of Computer Science, The George Washington University, where he has been teaching and performing research since 1980. Previously, he was employed at the Los Alamos National Laboratory in the Computer Graphics group where he worked in the general area of data visualization. He was educated at Wittenberg University, Miami University (Ohio) and received his Ph.D. from the University of Michigan. He was an associate editor of *ACM Transactions on Graphics* and is one of the co-founders of the *ACM*

Symposium on User Interface Software and Technology. He has received funding from NSF, NASA, USGS, FAA, DISA, NSA, DARPA, and IBM. His past research has included development of architectures to support highly interactive user interfaces, use of computer graphics to support visualization of large data sets, and object oriented design methodologies. He is currently working on novel interaction techniques including eye-tracking and vibrotactile feedback.

James K. Hahn is currently a full professor in the Department of Computer Science at the George Washington University where he has been a faculty since 1989. He is the founding director of the Institute for Biomedical Engineering and the Institute for Computer Graphics. His areas of interests are: medical simulation, image-guided surgery, medical informatics, visualization, and motion control. He received his Ph.D. in Computer and Information Science from the Ohio State University in 1989, an M.S. in Physics from the University of California, Los Angeles in 1981, and a B.S. in Physics and Mathematics from the University of South Carolina in 1979.