AniDance : Real-Time Dance Motion Synthesize to the Song

Taoran Tang*
Tsinghua University
Beijing, China
463618845@qq.com

Hanyang Mao*
Tsinghua University
Beijing, China
maohanyang789@163.com

Jia Jia†
Tsinghua University
Beijing, China
jjia@mail.tsinghua.edu.cn

ABSTRACT

In this paper, we present a demo named AniDance that can synthesize dance motions with melody in real-time. When users sing a song or play one in their phone to AniDance, their melody will drive the 3D-space character to dance to create a lively dance animation. In practice, we conduct a music oriented 3D-space dance motion dataset by capturing real dance performances, using LSTM-autoencoder to identify the relation between music and dance. Based on these technologies, users can create valid choreographies that capable of musical expression, witch can promote their learning ability and interest in dance and music.

CCS CONCEPTS

• **Human-centered computing** → *HCI theory, concepts and models*; • **Computing methodologies** → Supervised learning by regression;

KEYWORDS

Multisensory interaction, Motion Synthesis, LSTM, Autoencoder, Music-dance dataset, 3D motion capture

ACM Reference format:
Taoran Tang, Hanyang Mao, and Jia Jia. 2018. AniDance : Real-Time Dance Motion Synthesize to the Song. In *Proceedings of 2018 ACM Multimedia Conference, Seoul, Republic of Korea, October 22–26, 2018 (MM '18),* 3 pages.
https://doi.org/10.1145/3240508.3241388

1 INTRODUCTION

Human activities based on multisensory interaction can enhance the transmission of information and promote the development of human brain[5]. Music and dance are closely related, a multisensory interaction of dance creation that harmoniously utilize auditory, motor, and visual senses. For example, synthesizing dance motions by music will lead to promotion of people's learning ability and interest in dance and music. Although many researchers have tried, synthesizing music-oriented dance motions still face many challenges, namely: 1) How to achieve more efficient multisensory interaction; 2) How to choose appropriate dance motions and make artistic enhancements to the choreography according to music; 3) A lack of training data.

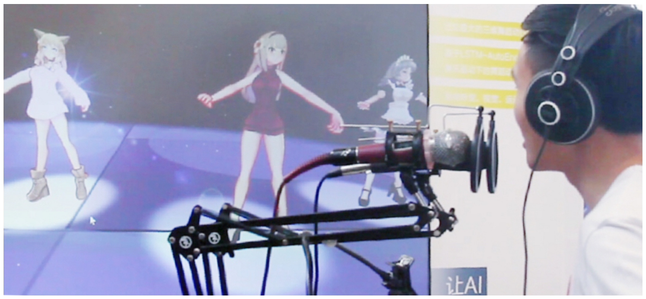

Figure 1: The AniDance system, the user stands in front of the screen, when he/she is singing to the microphone, the characters on the screen will dance with the voice.

In this paper, we will mainly consider how melody changes affect the way motions represented. In the real world, when the melody changes, the dancer does not only tend to change the motion itself, but also the motions' speed and intensity. Traditional dance synthesis algorithms randomly select dance motions from the database,therefore, the synthesized choreographies have little linguistic or emotional meanings. We promote a Real-Time Dance motion synthesis to the voice of singing based on LSTM-autoencoder named AniDance, which help users creating lively 3D-space dance animation by their voice of singing, as shown in Figure 1.

The contributions of this paper are summarized as follows: 1) We construct a music oriented 3D-space dance motion dataset using optical motion capture equipment (Vicon), which contains 40 complete dance choreographies in four types of dance. As far as we know, this is the largest music-dance dataset. We are willing to make our dataset open to facilitate other related research. 2) We propose a music-oriented dance synthesis to better understand and to identify inner patterns between acoustic features and motion features, so the emotion of the music will be reflected by the synthesized dance. This allows us to learn how people adjust their local joint posture and motion rhythm to express the changes in musical emotions and the rules of choosing motions in choreography.

2 DATASET

We asked professional dancers to dance with music, and captured their motions using optical motion capture equipment. Thus we construct a music oriented 3D-space dance motion dataset using optical motion capture equipment (Vicon), which contains 40 complete dance choreographies in four types of dance (Waltz, Tangle,

[*] These authors contributed equally to this work and should be considered co-first authors
[†] Corresponding author

MM '18, October 22–26, 2018, Seoul, Republic of Korea
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-5665-7/18/10…$15.00
https://doi.org/10.1145/3240508.3241388

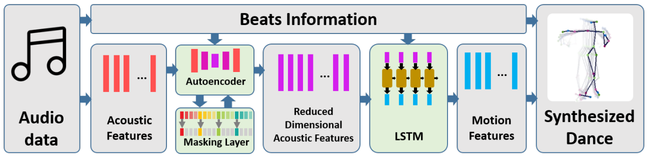

Figure 2: The structure of the LSTM-autoencoder networks.

Cha-Cha, and Rumba) totaling 907,200 frames and 94 minutes in general. As far as we know, this is the largest music-dance dataset. We are willing to make our dataset open to facilitate other related research. [1] The original data contained 41 joints in 3D space, and we manually picked out 21 joints best represent the dancers' motions as the *motion features* of the dancer. We apply smooth and normalization algorithms onto our dataset to make our data more regular.

3 TECHNOLOGY

The technical part of our application contains four modules, including real-time beat detection, acoustic features extraction, motion prediction, and dance exhibition.

Real time beats detection. While the dance motions must step on the beats, to generate the dance motions in time, it is necessary to predict the exact time of the forthcoming beat. Existing research provides offline beats detection [2]. On the basis of previous study, a real-time beats detection tool is established using several self-adaption algorithms.

Acoustic features extraction. An audio analysis library, named *librosa*, is used for music information retrieval as proposed by McFee [4]. We use the *Mel Frequency Cepstrum Coefficient (MFCC)*, *Constant-Q Chromagram*, *Tempogram*, *Onset strength* provided in librosa as the acoustic features of our network.

Motion prediction. The network we used for prediction is a combination of Autoencoder [1] and LSTM [3]. The structure of our network shows in figure 2. We transform the 21 joints of motions in our dataset into three dimensional vectors, which are used as motion features. And train the LSTM-autoencoder networks with acoustic features as input and motion features as output.

In LSTM-autoencoder networks, an autoencoder network reduces the dimension of acoustic features, and a masking layer was used to prevent over-fitting, the lower-dimensional acoustic features are fed into LSTM network, the dance motion can hence be synthesized in the form of positions of the skeleton nodes.

Based on the characteristics of out tasks. We apply several tricks to the network for better performance. Several acoustic features represent essential temporal information of the songs. We pick them out and feed them directly into the LSTM layers.

We use quantitative (euclidean distance) and qualitative experiments (user study) to evaluate the performance of our network. The network performs best in both quantitative and qualitative measurements is chosen to be used in practice.

Dance exhibition. Instructed by the predicted motion features, we implement skeleton animation on carton characters for real-time virtual dance exhibition.

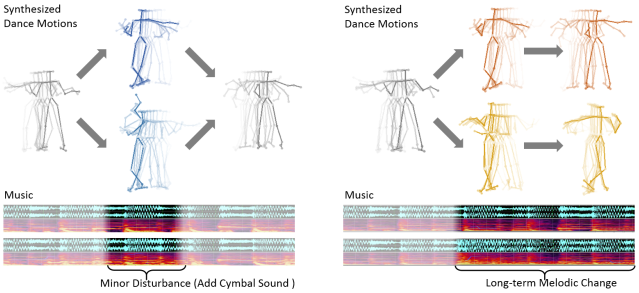
Synthesized Dance Motions
Music
Minor Disturbance (Add Cymbal Sound)
Synthesized Dance Motions
Music
Long-term Melodic Change

Figure 3: The dance motions synthesized by our model according to different changes of music. In this figure, minor changes in music will only affect the subtle adjustments of dance motions, while apparent changes in music will lead to entirely different motions.

4 DEMONSTRATION

In demonstration, we use the professional recording equipment to capture user's voice. Microphone, SDP Cara OK effector and closed back headphone creates unprecedented experience for the users, and the dance animation are shown on a large screen. Different operating modes are provided in our application. Users are free to choose the dance type, the dancer's speed, or even ask the dancers follow his/her speed of singing.

We tried to verify that AniDance could solve the challenges discussed in Section 1. First, we added a short period of drumbeat into a melody. It turned out the synthesized dance did change as invoked by the drumbeats, and when the drumbeats ended, the dance went back to it's original states. Then we tried to replace the melody with a completely new one, and the entire segment of choreography also changed. The results are shown in Figure 3, which proved that AniDance is effective and efficient in synthesizing valid choreographies which are also capable of musical expression.

5 CONCLUSIONS

In this paper, we present a demo named AniDance, based on a LSTM-autoencoder network, AniDance can synthesize dance motions with melody in real-time. We use acoustic features as input to get the output of synthesized dance choreographies with music which has a richer expression and better continuity. Our future work will focus on three major aspects: 1) Acquiring more data to continuously enhance our dataset. 2) Taking into consideration the users' different preferences about the synthesized dances, and 3)Leverage our model to build various applications.

More examples of dance motions synthesized by music are shown in the video of the supplementary material, and *Dance with Melody: An LSTM-autoencoder Approach on Music-oriented Dance Synthesis* that submitted to the full paper section of ACM Multimedia 2018.

6 ACKNOWLEDGMENTS

This work is supported by National Key Research and Development Plan (2016YFB1001200),the Innovation Method Fund of China (2016IM010200).

[1] https://github.com/Music-to-dance-motion-synthesis/dataset

REFERENCES

[1] Pierre Baldi. 2012. Autoencoders, unsupervised learning, and deep architectures. In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*. 37–49.

[2] Daniel PW Ellis. 2007. Beat tracking by dynamic programming. *Journal of New Music Research* 36, 1 (2007), 51–60.

[3] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[4] Brian McFee, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. 2015. librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference*. 18–25.

[5] F. H. Rauscher, G. L. Shaw, and K. N. Ky. 1995. Listening to Mozart enhances spatial-temporal reasoning: towards a neurophysiological basis. *Neuroscience Letters* 185, 1 (1995), 44–47.