AniDraw: When Music and Dance Meet Harmoniously

Yaohua Bu,[12†] Taoran Tang,[12†] Jia Jia,[1*] Zhiyuan Ma,[1] Songyao Wu,[2] Yuming You[1]
[1]Department of Computer Science and Technology, Tsinghua University, Beijing, China
Tsinghua National Laboratory for Information Science and Technology (TNList)
Key Laboratory of Pervasive Computing, Ministry of Education
[2]Academy of Arts & Design, Tsinghua University, Beijing, China
{byh15, ttr15}@mails.tsinghua.edu.cn, jjia@mail.tsinghua.edu.cn

Abstract

In this paper, we present a demo, AniDraw, which can help users practice the coordination between their hands, mouth and eyes by combing the elements of music, painting and dance. Users can sketch a cartoon character through multi-touch screens and then hum songs, which will drive the cartoon character to dance to create a lively animation. In technical realization, we apply the mechanism of acoustic driving in which AniDraw extracts time-domain acoustic features to map to the intensity of dances, frequency-domain ones to map to the style of dances, and high-level ones including onesets and tempos to map to start, duration and speed of dances. AniDraw can not only stimulate users' enthusiasm in artistic creation, but also enhance their esthetic ability on harmony.

Introduction

In art, music and dance are closely related, yet they are taught separately in traditional model of art education, which forms an invisible shackle of creativity. Therefore, we present an iPad-based application named AniDraw that can explore the potential relationship of different art forms when applied in art education. AniDraw supports users to draw cartoon characters at will and make them dance to their humming songs in real time.

Most audio-visual multimodal integration applications are not realized in real time (Raffle et al. 2007). Motivated by this kind of existing applications using audio driving animation (Levine et al. 2009), Anidraw improves in that it can take an arbitrary sketched character as input, and is tailored for music and dance (rather than conversational speech). We will demonstrate the contributions as: (1) Firstly, propose the method of semi-auto recognition with predefined skeleton model to locate carton characters joint even if they are not complete. (2) Secondly, use acoustic features to drive dance movement generation in real time to realize the harmony of music and dance.

Furthermore, AniDraw offers children the opportunity to explore and nurture their creative and artistic instincts, and strengthen their perceptions. For middle-aged and elderly users, AniDraw can also improve their mental and physical health, because it can not only help users to obtain a sense of achievement in artistic creation, but also enhance their rhythmicity and body coordination by practicing their eyes, hands and mouths together.

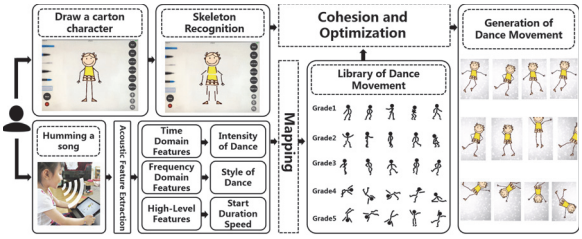

Figure 1: Technical realization framework of AniDraw.

Technical Realization

In Figure 1, we elaborate our framework of acoustic driving mechanism, which mainly involves three units including skeleton recognition, acoustic feature extraction and mapping to dance, and our own library of dance movement. Details are stated in the following sections.

Skeleton Recognition

We use skeleton animation to represent characters drawn by users, which connects all the nodes of key body parts into a tree structure, and construct a set of affinity transformation to represent the action space. In order to take better segments of characters, we separate the cartoon character into predefined parts including the skin and 13 typical body parts.

Acoustic Feature Extraction

Acoustic driving means to extract acoustic features of user's songs in real time and based on these to determine the action at next time-step. We extract time-domain features including loudness and pitch, frequency-domain features including Fourier Transform Spectrum and MFCC, and high-level features including onsets and tempos. We use YIN-FFT algorithm (Brossier et al. 2006) to detect base frequency. And for detecting the Starting Point of Notes in complex fields, we take a robust algorithm (Duxbury et al. 2003) which combines the energy and phrase.

*Corresponding author: J.Jia (jjia@mail.tsinghua.edu.cn)
†Joint first authors: Yaohua Bu, Taoran Tang (According to the last alphabetical order)

Mapping to Dance

After extracting the multidimensional acoustics features stated above, we design a model that can take adaptive selection from our own library of dance movements. (To establish this library, we record 24 kinds of different styles of classical dance, and divide them into five different levels of intensity according to the range of motions and the expression of actions.)

- **Determine the Start, duration and Speed of the Dance** We analyse the onset and tempos of songs to determine the start, duration and speed of the dance. To deal with the relative error between the prediction and our ground truth, we take methods including continuation of the end of the action, early termination, and loop playback to realize better fit. Apart from these, we apply an event-driven mechanism at the end of current action, which request to start the next cycle of selection by updating a new threshold each time-step.
- **Determine the Style and Intensity of the Dance** We estimate the identity of singers through use the MFCC to choose a suitable style of dance movement. We record the loudness and pitch at the beginning of each note, and determine the intensity of the dance movement by dynamically setting the threshold to control the intensity of the current moment relative to the previous singing.
- **The Optimization for the Cohesion of the Dance** At each moment, AniDraw takes synthesis of the skeleton recognized from users' real-time drawing and the sketch selected from the library of dance movement to complete dance generation. In order to make the connection between two movements smoother, we establish some strategy for optimization including, to control each moving distance, to avoid characters moving out of the picture, to set a threshold which can reduce the probability of the same action to be chosen repeatedly at neighborhood times.

User Study

Our evaluation of AniDraw require to reply the following two research points: (1) Whether AniDraw is user-friendly enough; (2) How effective AniDraw can help to improve users' enthusiasm in artistic creation.

Design We set up two experimental groups including the supervised group (signed as G1) in which participants would be taught how to operate AniDraw, and the unsupervised group (signed as G2) in which participants explore AniDraw on their own. We invited 52 children (26 females, 26males) from a kindergarten, and then divided them into two groups with the same gender distribution. Apart from this, we designed a quantitative measurement consisting of three variables as in table 1. And we estimated their values by observing and recording emotional changes, language feedback and action behavior of all participants.

Results The results of qualitative comparison between G1 and G2 indicate that AniDraw is easy to operate and user-friendly enough. Furthermore, based on the value of t, φ_1 and φ_2 from both two groups, we can find that Anidraw helps participants focus on their creation and brings them great happiness. This proves perfectly that Anidraw can stimulate much enthusiasm of users artistic creation.

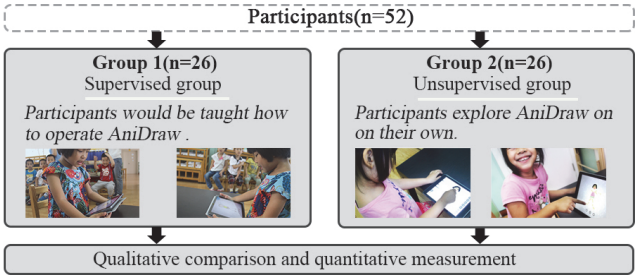

Figure 2: Qualitative comparison and quantitative measurement.

Table 1: Variables for the quantitative measurement

Notation	Variable Declaration
t	how long participants focus in their own creation of cartoons
φ_1	how many times participants smile
φ_2	the degree from 0 to 5 about whether participants like AniDraw

Conclusions

Based on acoustic driving mechanism, Anidraw can help users, especially children, explore principles of harmonious beauty involving in colors, shapes, music and dance, which will truly inspire their great enthusiasm to participate in artistic creation. In the future, we plan to apply some DNN-based models to generate a better and richer library of dance movements for Anidraw.

Acknowledgement

This work is supported by the National Key Research and Development Plan (2016YFB1001200, 2016IM010200), and National Natural and Science Foundation of China (61370023, 61602033).

References

Raffle, H., Vaucelle, C., Wang, R. and Ishii, H. 2007. Jabberstamp: embedding sound and voice in traditional drawings. In *Proceedings of the ACM SIGGRAPH on Educators Program*, NY, US.

Levine, S., Theobalt, C., Koltun, V. 2009. Real-time prosody-driven of body language. In *Proceedings of the ACM SIGGRAPH Asia on Transactions on Graphics*, NY, US.

Duxbury, and Chris 2003. Complex domain onset detection for musical signals. In *International Conference on Digital Audio Effects Workshop*, London, UK.

Brossier, and Paul, M. 2006. Automatic annotation of musical audio for interactive applications, Department of Digital Music, Queen Mary University of London, London, UK.