Automatic Facial Animation Generation System of Dancing Characters Considering Emotion in Dance and Music

Wakana Asahina[1, 2*] Narumi Okada[1, 2] Naoya Iwamoto[1, 2] Taro Masuda[1, 2] Tsukasa Fukusato[1] Shigeo Morishima[2, 3†]

[1] Waseda University [2] JST CREST [3] Waseda Research Institute for Science and Engineering

1 Introduction

In recent years, a lot of 3D character dance animation movies are created by amateur users using 3DCG animation editing tools (e.g. MikuMikuDance). Whereas, most of them are created manually. Then automatic facial animation system for dancing character will be useful to create dance movies and visualize impressions effectively. Therefore, we address the challenging theme to estimate dancing character's emotions (we call "dance emotion"). In previous work considering music features, DiPaola et al. [2006] proposed music-driven emotionally expressive face system. To detect the mood of the input music, they used a hierarchical framework (Thayer model), and achieved to generate facial animation that matches music emotion. However, their model can't express subtleties of emotion between two emotions because input music divided into few moods sharply using Gaussian mixture model. In addition, they decide more detailed moods based on the psychological rules that uses score information, so they requires MIDI data. In this paper, we propose "dance emotion model" to visualize dancing character's emotion as facial expression. Our model is built by the coordinate information frame by frame on the emotional space through perceptional experiment using music and dance motion database without MIDI data. Moreover, by considering the displacement on the emotional space, we can express not only a certain emotion but also subtleties of emotions. As the result, our system got a higher accuracy comparing with the previous work. We can create the facial expression result soon by inputting audio data and synchronized motion. It is shown the utility through the comparison with previous work in Figure 1.

2 Building of Dance Emotion Model

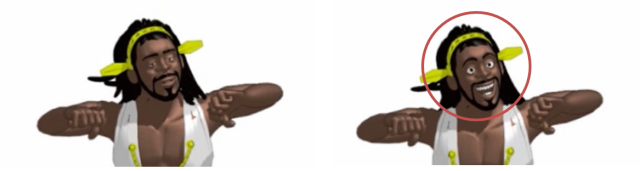
Our system consists of three parts: subjective evaluation experiment, parameters detection, and facial animation, for building a dance emotion model.

Mood decision based on Subjective Evaluation Experiment:

To build a dance emotion model, we performed a subjective evaluation experiment and got impression evaluation values annotated every second for dance movies. We used 400 dance movies which had been made by combining four different types of dance motions (high-active, low-active, large-shape, small-shape) with 100 pop music clips from the RWC music database. Then, each of 10 subjects evaluated the impression of those movies every second on the V-A plane (an emotional space) that consists of two fundamental dimensions; valence and arousal.

*e-mail: 2236-wakana@fuji.waseda.jp

†e-mail:shigeo@waseda.jp



(a)Thayer model (b)Our model

Figure 1: *One example of facial animation. The impression of input dance movie is high 'arousal'. Therefore "exciting" is chosen for the character's facial expression through our method (b).*

Parameters for Dance Emotion Model:

We used multiple regression analysis to build a model and chose both music and motion features as parameters for it. We chose 13 dimensional vectors as motion features based on Laban movement analysis that is a theory that defines the relationship between a varieties of body movements and mental states. We also chose 80 dimensional vectors as the music features that are related to rhythm and timbre and that only need audio signals based on previous work.

Generate Facial Animation Based on the V-A plane:

We express all facial animation based on the V-A plane. Based on Russell's basic emotions, we divided the V-A plane into eight equal parts and chose as many emotional states. Concretely, "arousal", "excitement", "pleasure", "relaxation", "sleepiness", "depression", "displeasure" and "distress" are used as basic facial emotions. All facial animations on the V-A plane can be expressed by using a linear combination of the two adjacent emotional faces.

3 Results and Conclusions

We compared the impression estimation accuracy of our method with that of Thayer model. The result shows that the precision rate is improved by using our method because we also considered motion features (Table 1). Besides there seemed a high correlation between the trajectories of the V-A coordinates and that of the motion feature values over time. Moreover in our system, we can express any subtle facial animation on the V-A plane, which indicates that we can generate more natural facial animations than those generated by the previous work. In addition, considering that we need only audio signals and no MIDI data, the versatility of our system is higher than that of other systems that need MIDI data.

In conclusion, we proposed the automatic generation system of facial animation of dancing characters by considering music features and motion features. For future work, we plan to clarify the relationships between music, emotions and dance motions.

Table 1: *Precision rate / Coefficient of determination (r2)*

	Precision rate	r2(Valence)	r2(Arousal)
Our method	0.6500	0.7178	0.3788
Thayer model	0.2958	-	-

References

DiPaola, S., and Arya, A. 2006. Emotional remapping of music to facial animation. In *Proceedings of the 2006 ACM SIGGRAPH Symposium on Videogames*, ACM, New York, NY, USA, Sandbox '06, 143–149.