<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *GCCE2017*.

Citation for the original published paper:

Hayashi, M., Bachelder, S., Nakajima, M., Shishikui, Y. (2017)

Rap Music Video Generator: Write a Script to Make Your Rap Music Video with Synthesized Voice and CG Animation.

In:

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-336338>

Rap Music Video Generator

Write a Script to Make Your Rap Music Video with Synthesized Voice and CG Animation

Masaki Hayashi, Steven Bachelder, Masayuki Nakajima
Department of Game Design
Uppsala University
Visby, Sweden
masaki.hayashi@speldesign.uu.se

Yoshiaki Shishikui
Department of Frontier Media Science
Meiji University
Tokyo, Japan
sisikui@meiji.ac.jp

Abstract— We have made an application to make rap music video with CG animation by writing out a simple script. Aquestalk and TVML (TV program Making Language) are used for synthesized voice and real-time CG generation, respectively. A user can enjoy making rap music video easily by writing speech texts and character movements along with the music beat in the script.

Keywords— CG animation, Voice synthesis, Music, TVML, Aquestalk

I. INTRODUCTION

Rap music was born originally from American hip-hop culture, but spread to the world very quickly, and now it is becoming a definite singing style. In many countries, the rap music performed in each domestic language has spread rapidly. We have developed a rap music generator from text typed in by a user. The application is based on the concept of converting text to animation using CG (Computer Graphics) characters speaking with synthesized voices [1]. Our application also covers not only the rap music section with singing the input text on the music beat, but also ordinary acting sections of the same character such as talking to a microphone with gesture and walking away or whatever on a stage. This means that the user can make a music video of full stage acting with rap music easily by typing in a script on a text editor.

II. METHODS OF MAKING RAP MUSIC VIDEO

The idea is: CG animation of rap music is generated by a user script. We first setup a language specification of the rap script, then the specification is implemented to the application. Currently, only Japanese version is developed but it could be extended to any languages with the same method theoretically.

For the synthesized speech, we use Aquestalk2 of AQUEST Corp. (<http://www.a-quest.com/>) and TVML (TV program Making Language) for CG character animation [2]. The Aquestalk2 is a lightweight speech synthesis engine, and it is widely used in Japan because it is open to the public for non-profitable use. Also, the Aquestalk2 opens the specification of "phonetic symbols" so that a user can designate fine tuning of speech accent, which is indispensable for increasing the degree of completion of rap singing. On the other hand, the TVML is a technology developed by the author, which enables to make

television-program-like CG animation from a script written in the TVML language specification. The TVML playback engine is provided as a form of SDK (System Development Kit), and that is used to build the application.

III. DETAIL DESCRIPTION OF THE APPLICATION

A. The script

The basis of the specification of the rap music script is a combination of the position of the beat of the rhythm of music and the rap text. For example:

2-4: バッサバッサとなぎ倒す

The syntax is "bar number-beat number: text". In the above example, the speech text "バッサバッサとなぎ倒す" (pronounces "Bassa bassa to nagi taosu") starts uttering from the 4th beat in the 2nd bar. Bars and beats start from 1 (not from 0), and in the current implementation, the beat resolution is 16th note (therefore, the beat numbers vary from 1 to 16).

The speed of uttering can be altered by giving an auxiliary command as follows:

4-1: 途中で倒れて死んでも(speed=2)
4-9: いいんじゃないーい？

With the two lines above, the utterance "途中で倒れて死んでも" ("Tochuu de taorete shindemo") with "double speed" from the head (1st) of the 4th bar, and the next line "いいんじゃないーい？" (Iinja naai?) starts at the 9th beat in the same bar at normal speed. A line with no speed designation in parentheses is uttered at the default speed (means "speed=1"), which is a preset to the utterance speed of the Aquestalk2 reading 16 syllables within a bar. This means that one syllable corresponds to a sixteenth-note in a bar. (e.g. "あなたの事など聞きたくないから" ("A-n-ta-no-ko-to-na-do-ki-ki-ta-ku-na-i-ka-ra") occupies a full bar)

If you want to use the phonetic symbol notation of Aquestalk2, a series of symbols should be surrounded by <at> tag.

```
1-2:<at>あ'い'は'い'づ</at>(speed=1.8)
1-5:<at>べ'ん</at>
```

By using this notation, it is possible to specify fairly detailed accents. Also, in this case, subtitles that appear automatically would appear as they are with the series of symbols, so the "subtitle notation" can be used to designate the subtitles separately as follows.

```
1-1: (subtitle= I have a PEN)
1-2:<at>あ'い'は'い'づ</at>(speed=1.8)
1-5:<at>べ'ん</at>
2-1: (subtitle off)
```

In addition to the rap music script, ordinary acting commands is available. When you use the rap section and the ordinary script by using "(start of rap)" and "(end of rap)" as follow:

```
Narrator: Next one, please.
Hello, my name is Bob. Hit it!
(start of rap)
1-1: (subtitle= I have a PEN)
1-2:<at>あ'い'は'い'づ</at>(speed=1.8)
1-5:<at>べ'ん</at>
2-1: (subtitle off)
(end of rap)
How was it? Do you like it?
Narrator: No. You have failed.
```

The section which is not surrounded by those "start" and "end" is regarded as an ordinary acting.

B. The Animation

The TVML engine generates CG animation in real-time. The engine has an automatic lip sync function, so that you do not need to deal with the lip sync. The dance of the character synchronized with the music, is done by giving a TVML pose command to the character in accordance with the beat. The camera angle is automatically switched among the angles prepared in advance. You can also explicitly designate your desired camera angle in a script.

C. The Application

The application is built on the Unity game engine (Version 5.4.4). It has only play / stop buttons and a script file selection button. The playback script is an ordinary text file, and the script is edited with a normal text editor. The application is currently Windows only however basically it is relatively easy to support other platforms since Unity is a multi-platform software. Also we plan to support other languages such as English. The biggest concern is the Text-To-Speech engine of the language. The current Japanese TTS engine used in the application is working in real-time because of its very small utterance latency due to a lightweight engine design. When we support other languages, the same type of low latency TTS would be needed or we would take off-line calculation of the TTS engine which may require much work. However, theoretically, there is no problem for making other language version.

Figure 1. Screenshots of generated rap music vide

Fig. 1 shows some screenshots of the music video created by our application. You can watch the entire video at: <https://youtu.be/urh7DWshIxI>. You can see the entire script put on the video description in the YouTube entry. And, the download site of a test version of the application is here: http://tvmlab.com/?page_id=148.

IV. CONCLUSION

We have developed an application that makes a rap music video just by writing a script. Even in the world of Hatsune-Miku by Vocaloid, rap-like productions are actively done, however, it still requires a high level of skill to create such a thing. Since this application is very easy-to-use, only to write a script and to press the play button, there is no need to become proficient of the complicated graphical user-interface operation. It is so simple to use, but of course you need to master how to write a script.

We plan to enhance the application so that users can enjoy more fun by allowing them to freely select and register multiple characters, voice types, dance patterns, rap music rhythm tracks in the future.

REFERENCES

- [1] M. Hayashi, S. Inoue, M. Douke, N. Hamaguchi, H. Kaneko, S. Bachelder, M. Nakajima: "T2V: New Technology of Converting Text to CG Animation", ITE Transactions on Media Technology and Applications, Vol.2, No.1, pp.74-82 (2014)
- [2] M.Hayashi, "TVML (TV program Making Language) - Automatic TV Program Generation from Text-based Script -", Proceedings of Imagina'99, pp.119-133 (1999)