REAL-TIME FUTURE-RHYTHM VISUALIZER FOR DJ PERFORMANCE

Masatoshi HAMANAKA (masatoshi.hamanaka@riken.jp) (0000-0002-0249-4900)[1]

[1]**RIKEN**, Tokyo, Japan

ABSTRACT

In breaking, which features two dancers breakdancing against each other to the rhythm of music remixed by a disc jockey (DJ), it is important to match the technique to the timing of the music beat generated by a DJ. However, it is difficult for many in an audience to judge whether the timing is correct. We developed a real-time future-rhythm visualizer that superimposes in real time visualized video of beats from the past to the future of the music remixed by a DJ on the live broadcast video of breaking. However, the timing of the two does not match due to transmission and equipment delays when the video camera is capturing video and the computer is playing visualization video. Therefore, we made it possible to adjust the timing by changing the position of the object representing the present time. The visualizer was used at the All-Japan Breaking National Championships recorded by NHK, a national TV station in Japan.

1. INTRODUCTION

Breaking, an official competition of the Paris Olympics, features two dancers breakdancing against each other to the rhythm of music remixed by a disc jockey (DJ). The DJ uses two turntables to play multiple musical pieces while switching between them. The DJ then adds accents to the music by scratching the turntable.

Freezing while performing tricks, such as inverted stands, in perfect time with the rhythm generated by the DJ is called "kill the beat" and is the highlight of the competition, as it requires a high level of skill. Most of the audience at a breaking venue have experienced breaking, and they become very excited when they see a successful kill the beat.

However, viewers who are watching breaking for the first time have trouble judging the success of a kill the beat on TV or video. Therefore, we developed a real-time future-rhythm visualizer that displays the future rhythm of the music being remixed by the DJ.

Several methods were proposed to automatically generate a setlist for DJs to prepare before remixing [1–3]. An automated DJ-mixing system was also proposed [4] as well as a method for smoothly crossfading from one piece to another [5].

While these studies [1–5] intended to automate the DJ's work or assist the DJ, our study is intended to assist the breaking audience through visualization. Rhythm visualization helps the audience understand if the rhythm matches the timing of the moves. Visualization will also enable the hearing impaired to understand the timing of the rhythm and enjoy breaking.

Our visualizer was used on February 2024 at the All Japan Breaking Championships recorded by the Japan Broadcasting Corporation (NHK), a national TV station in Japan (Fig. 1). The recorded video can be viewed on NHK's website.

Figure 1. Live recording by NHK of our real-time future-rhythm visualizer

2. VISUALIZATION OF FUTURE RHYTHMS

In breaking, dancers improvise to the pieces remixed by the DJ, so it is not disclosed in advance which pieces will be used. Many pieces used by DJs include local tempo change because they make new pieces remixed from old pieces. The tempo of a piece can be changed by operating the DJ mixer. The DJ can also manipulate the disc to temporarily stop a piece or stretch or shorten a sound. Therefore, it is difficult to visualize future rhythms in music remixed by DJs.

2.1 Future Rhythms are Unpredictable

It is difficult to predict future rhythms of music remixed by DJs because changes are added to the music as the disc is manipulated. Therefore, we receive all pieces used by DJs, analyze them, and generate rhythm-visualization source videos. The sound and video are played back in sync by creating a link between the pieces and source video on the DJ software (Fig. 2) [6]. The rhythm-visualization source video consists of long bars representing down beats and short bars representing up beats, which scroll vertically depending on time. The interval between the bars is constant, and the scrolling speed varies depending on the tempo. When the DJ increases the playback tempo of the music, the playback speed of the video also increases; thus, the speed at which the bars scroll vertically increases.

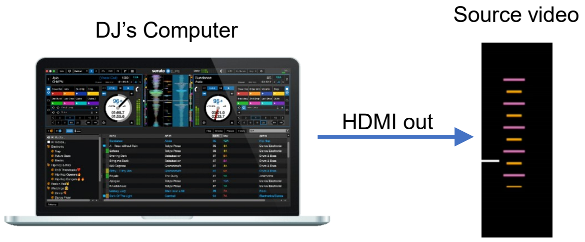

Figure 2. Sound and video are played back in sync

2.2 Appropriate Musical Analysis

The up and down beats can be obtained through artificial intelligence (AI) analysis. However, simple analysis by the AI sometimes results in the up and down beats being reversed, depending on the timbre of the instrument or other factors. Therefore, we added a musical analysis to our visualizer to solve the problem of up and down beats being reversed [7–10].

3. REAL-TIME AND NON-REAL-TIME RHYTHM VISUALIZATION

The broadcast video produced by NHK includes delays as it passes through selectors that integrate multiple cameras and switchers that superimpose information, etc. The distance between the cameras and broadcast room at NHK Hall, the venue of the Championships, was more than 100 meters, which causes transmission delays. Our visualizer also needs to send information from the DJ booth to the broadcast room, which causes delays.

3.1 Adjust Current Timing Indicator

Due to these delays, the timing between the dance video by NHK and the rhythm-visualization video from the DJ's computer is shifted. Therefore, we can adjust the position of the rhythm-visualization output video by adding a line indicating the current time on the source video later in the program so that the timing of the dance video and the rhythm-visualization output video are aligned (Fig. 3).

3.2 Animation for Current Timing

Adding an effect, such as a color change or animation, when the up or down beat bars pass the current timing makes the timing easier for the viewer to understand. However, a video with effects added in advance cannot be created because the timing and position change depending on the delay in the previous item. Therefore, we developed a program for our visualizer that analyzes the rhythm-visualization video being played, reads the bar positions, and adds effects at the appropriate time.

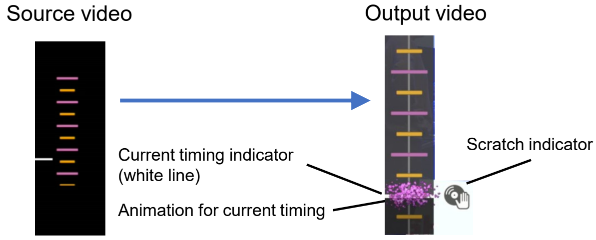

Figure 3. Sound and output video

4. SCRATCHING VISUALIZATION

Two turntables are connected to the DJ mixer, and the DJ software can be used to set the pieces assigned to each turntable. When using the DJ software, a vinyl record, in which sine waves with different phases are recorded on the left and right sides of the vinyl record, is placed on the turntable. The DJ software receives the rotation information of the turntable from the digital vinyl system (DVS) signal generated by the vinyl record. The DJ software then receives the turntable rotation information from the generated DVS signal.

4.1 Detection of Turntable Rotation

Because DVS signals are analog signals, we branch them out, receive them at an audio interface, and analyze them to determine the rotation per minute of both the left and right turntables. Specifically, accelerating the turntable increases the frequency, and decelerating the turntable decreases the frequency. Because the DVS signal is a stereo signal and is out of phase between left and right, it can also detect whether the rotation is forward or reverse depending on which waveform arrives first.

4.2 Detection of Crossfader

DJ mixers have a crossfader, which is used for a swap between two pieces or for adding scratching sounds. Scratching is a technique in which a turntable is operated to rub the other turntable during the playback of a certain piece to emphasize the rhythm with the sound generated. The crossfader can then be moved slightly to the turntable on the side that is being scratched to mix both the playback and scratch sounds.

Many DJs are constantly performing scratching operations on turntables with the crossfader closed and no sound coming out, and they open the crossfader to output the scratching sound when necessary timing. Therefore, to detect scratches, we need to detect not only the rotation of the turntable but also the position of the crossfader.

Many models of DJ mixers that can be connected to a computer via a universal serial bus (USB) cable use Musical Instrument Digital Interface (MIDI) signals for communication. Some DJ mixer models that are equipped with two USB ports for DJ alternation can receive the USB signal flowing through the port in use on the other USB port. Therefore, the MIDI signal is analyzed, and the crossfader position is extracted.

When a scratch is detected, a scratch indicator is displayed. Since the fader moving distance and the rotation speed at the time of a scratch vary greatly from DJ to DJ, the threshold for judging a scratch is manually changed for each DJ.

5. VISUALIZATION LAYOUT AND COLORS

The visualization layout was determined with reference to the NHK broadcast video from the previous year's competition. The entire screen is 1080 pixels in height and 2048 pixels in width, as 2024's broadcast was planned in 2K resolution. The dancers were displayed in the center of the screen, and various displays, such as program names, athlete names, and judgment results, were superimposed on the top and bottom of the screen. Therefore, it was decided to use a vertical display on the left edge so as not to interfere with the displayed dancers.

Of the 550 vertical pixels, approximately 440 are used to display rhythms from the future to the present, and the remainder are used to display past rhythms. The current timing indicator is displayed at the current position but can be moved up 20 pixels and down 20 pixels to absorb the delay. The current portion from the future is brightened and the past portion is darkened to draw attention to the future rhythm. The bar spacing was designed so that four down beats and four up beats are displayed (Fig. 4).

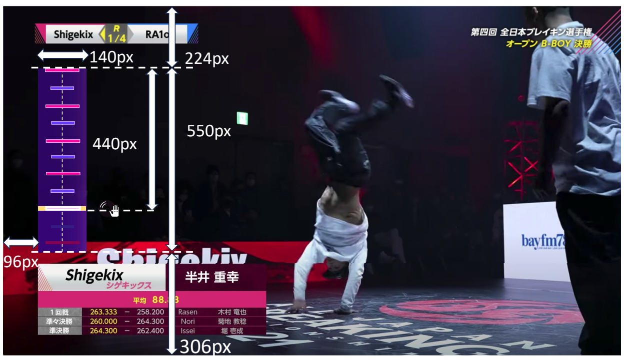

Figure 4. Layout of visualization area

The colors of the bars were selected with reference to the Tokyo Metropolitan Government's universal design guidelines [11], so that even people with diverse color vision can easily distinguish between down beat and up beat by color.

6. GENERATION OF VISUALIZATION SOURCE VIDEO

We received 300 copyright-free pieces from DJs performing at the competition, and generated visualization source videos for all of them.

6.1 Beat-timing Data

Music acoustic signals were analyzed to generate beat-timing data. Since many of the pieces had fluctuating tempos, tempo detection was executed from the audio data using the music-production software Logic Pro's Smart Tempo. For pieces with a beat per minute (BPM) higher than 130, the BPM was multiplied by 0.5. For pieces with a BPM of 64 or less, the BPM was set to 2 times. This is to prevent large changes, such as doubling or halving the speed of the bars in the visualization results, when the DJ smoothly connects from one piece to another.

MIDI messages representing the down-beat and upbeat were then written to each tempo-adjusted measure. When Logic Pro's Smart Tempo analysis did not produce the expected results, manual tempo adjustments were made.

6.2 Visualization Source Video Recorder

Beat-timing data are played back in Logic Pro to output MIDI messages representing the up and down beats and MIDI timing codes. We built a system to read those signals and record visualization source videos.

Figure 1 shows the system configuration. The hardware consists of a combination of non-special equipment. The visualization source video generated by computer (Mac Studio) is output via High-Definition Multimedia Interface (HDMI), which is then recorded by an HDMI capture device connected to the computer via USB-C.

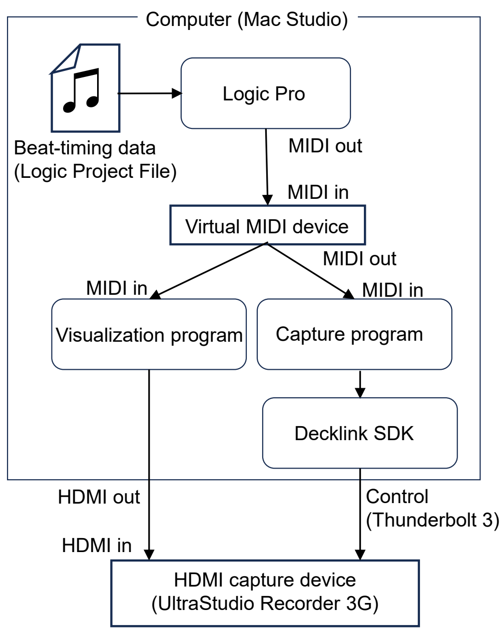

Figure 5. Visualization-source-video recorder

MIDI messages representing the up and down beats played by Logic Pro and MIDI timing codes are sent to the capture and visualization programs we built via a virtual MIDI device, a standard function of Mac OS.

The visualization program displays a bar on the screen when it receives a MIDI message indicating the up or down beat and changes the scrolling speed on the basis of the BPM information received by the MIDI timing code. In other words, the interval between bars is constant and the scrolling speed changes with the BPM.

The capture program starts recording when it receives a MIDI message and stops recording when no more MIDI messages are received.

We tried several HDMI capture devices, but those that execute hardware encoding when capturing video, such as Blackmagic Design's UltraStudio 4K Extream, do not have a constant delay between receiving the start of recording signal and actual start of recording, making it difficult to synchronize the start timing of the piece and video. However, Blackmagic Design's UltraStudio Recorder 3G, which encodes in software, had almost constant latency, so we decided to use it. The capture devices were controlled using the Decklink software development kit (SDK) from Blackmagic Design [12].

All the visualization source videos were generated in batch processing using Automatar, a standard function of Mac OS.

7. IMPLEMENTATION

Figure 6 shows an overview of our visualizer. We aimed to achieve the visualizer by combining general-purpose equipment without using special equipment. Turntables rotation information is obtained by branching the audio cable that connects the turntable to the mixer and by inputting it to the audio interface connected by a USB cable to a computer (Mac mini). The USB cable is connected between the mixer and Mac mini to acquire the crossfader position. The Mac mini then detects the scratch from the rotation and crossfader information, which is then transmitted to the computer (Mac Studio) via MIDI signals.

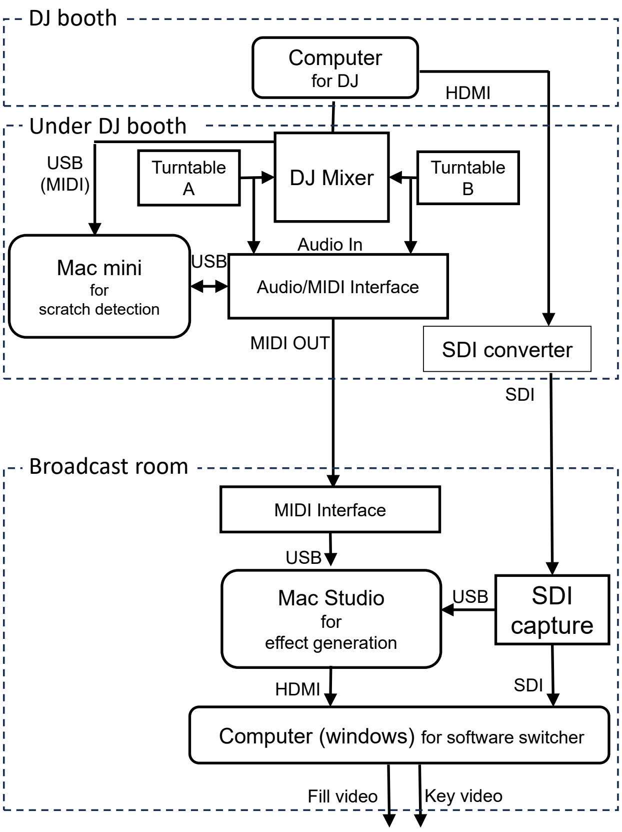

Figure 6. Overview of our real-time future rhythm visualizer

The video of the rhythm-visualization output from the computer used by a DJ is captured with Mac Studio, and image analysis is executed to detect when the bar passes the pixel set as current. Mac Studio then draws the scratch and animation of the bar as it passes.

Finally, the video drawn with Mac Studio and the rhythm-visualization video are captured with a Windows machine to generate the fill video and key video for the video composition.

The distance from the DJ booth to the broadcast room in NHK Hall (Fig. 7) was more than 100 meters, the details of the line between the booth and broadcast room are omitted from Fig. 6. For serial digital interface (SDI) transmission, we used the optical transmission equipment provided by NHK Hall, but it was SDI (3G-SDI) up to HD resolution. Since the output of the SDI converter was 4K-SDI (12G SDI), it was converted from 12G to 3G using an SDI down converter before transmission. MIDI transmission was executed by connecting an small form-factor pluggable (SFP) transceiver to a standard LEMO [13] fiber optic cable for the TV camera to construct a 10G-base network and using Network MIDI, a standard Mac function.

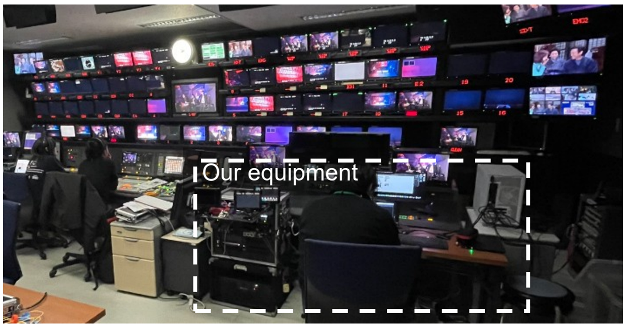

Figure 7. Our equipment in broadcast room in NHK Hall

8. EXPERIMENTAL RESULTS

Our visualizer was used at the Japan Breaking Championships in February 2024. Since the superimposed visualization video will be recorded live, we carefully checked the visualization source video in advance. Absorption of transmission delays was adjusted on site as it depended on the hall's equipment.

8.1 Video Screening

Using the system described in a previous study 6.2, 300 visualization source videos were exported in batch processing. The exported videos were manually screened, and half had problems. There were two causes of these problems: a timing error at the start of recording and image distortion. For the timing error, there was a blank space at the beginning of the video or the video started in the middle. For image distortion, the color of the bars was not uniform and there were unnecessary lines. After repeated exporting of the images with abnormalities and manual screening, all images with no abnormalities were generated after the eighth export.

8.2 Timing Screening

Although the timing of the exported visualization source video should be correct in principle, it was manually verified: 10 arbitrary pieces were selected out of the 300 pieces, and the visualization source video of the piece was tacked onto Logic Project and visually verified by a music and video expert. We confirmed that the timing was correct for all ten pieces, even after slow playback. For ten different pieces, the DJ software Serato [6] was used to create a link between the song and video, and the timing was visually checked while changing the BPM.

8.3 Absorption of Transmission and Equipment Delays

The video received from NHK was switched from multiple camera synchronized by time code. However, the sound on stage, the emcee's voice on stage, and the sounds remixed by the DJ were recorded by microphones. Synchronization of the video and sound was lip-synced by NHK using the voice and video of the emcee during rehearsals. The video was recorded after the rhythm-visualization video was superimposed on that video. When the recorded video was checked during slow playback, it was found that the visualization video was delayed by two frames at 59.94 frames per second. We then moved the current timing indicator down 20 pixels and re-recorded it; therefore, the visualized video and sound from NHK were synchronized.

9. CONCLUSIONS

We developed a real-time future-rhythm visualizer to visualize the music remixed by DJs in real time. The visualizer was used at the All-Japan Breaking National Championships and recorded by NHK, a national TV station in Japan. The recorded video can be viewed on NHK's website.

`https://www.nhk.jp/p/ts/9WXX8RKJVY/movie/` (access from Japan only).

When 20 people who viewed the recorded video were asked for their opinions, 18 said they enjoyed watching it, and 14 said "Kill the beat" was easy to understand.

The data generation and screening process involved a large amount of manual work, which we plan to pipeline and automate. In visualization-source-video export, there is currently a trade-off between timing error and image distortion. HDMI capture with software encoding has fewer timing errors but more image distortion. However, HDMI capture with hardware encoding has more timing errors but less image distortion. We plan to investigate ways to reduce both types of errors by using multiple machines and conduct more detailed experiments.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers 24H00748 and 21H03572.

10. REFERENCES

[1] R. M. Bittner, M. Gu, G. Hernandez, E. J. Humphrey, T. Jehan, H. McCurry, and N. Montecchio, "Automatic playlist sequencing and transitions." in *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR2017)*, 2017, pp. 442–448.

[2] T. Hirai, H. Doi, and S. Morishima, "Musicmixer: Automatic dj system considering beat and latent topic similarity," in *Proceedings of 22nd International Conference of MultiMedia Modeling (MMM2016)*. Cham: Springer International Publishing, 2016, pp. 698–709.

[3] J. Parera, "Dj codo nudo: a novel method for seamless transition between songs for electronic music," *Master's thesis, Universitat PompeuFabra, Barcelona*, 2016.

[4] Y. T. Hiromi Ishizaki, Keiichiro Hoashi, "Full-automatic dj mixing system with optimal tempo adjustment based on measurement function of user discomfort," in *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR2009)*, 2009, pp. 135–140.

[5] K. Robinson and D. Brown, "Automated time-frequency domain audio crossfades using graph cuts," in *Proceedings of the 20th International Society for Music Information Retrieval Conference (ISMIR2019) Late-Breaking/Demo*, 2019.

[6] Serato, "Serato DJ Pro - Professional DJ Software - Download — serato.com," https://serato.com/dj/pro, [Accessed 01-03-2024].

[7] M. Hamanaka, K. Hirata, and S. Tojo, "Implementing a generative theory of tonal music," *Journal of New Music Research*, vol. 35, no. 4, pp. 249–277, 2006. [Online]. Available: https://doi.org/10.1080/09298210701563238

[8] M. Hamanaka, K. Hirata, and S. Tojo, "Automatic generation of metrical structure based on gttm," in *Proceedings of the 2005 International Computer Music Conference (ICMC2005)*, 2005, pp. 53–56.

[9] M. Hamanaka, K. Hirata, and S. Tojo, "deepGTTM-I&II: Local Boundary and Metrical Structure Analyzer Based on Deep Learning Technique," in *Computer Music Modeling and Retrieval*, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:42476381

[10] M. Hamanaka, K. Hirata, and S. Tojo, "DeepGTTM-II: Automatic Generation of Metrical Structure Based on Deep Learning Technique." Proceedings of 13th Sound and Music Computing Conference (SMC2016), Aug. 2017. [Online]. Available: https://doi.org/10.5281/zenodo.851229

[11] Tokyo Metropolitan Government Bureau of Social Bureau of Social Welfare, “Tokyo metropolitan government color universal design guidelines,” https://www.fukushi.metro.tokyo.lg.jp/kiban/machizukuri/kanren/, [Accessed 01-03-2024](in Japanese).

[12] Blackmagic Design, “DeckLink ; Software | Blackmagic Design — blackmagicdesign.com,” https://www.blackmagicdesign.com/products/decklink/software, 2013, [Accessed 02-03-2024].

[13] LEMO, “About us | LEMO Connectors and cables — lemo.com,” https://www.lemo.com/our-world/company/about-us, [Accessed 07-03-2024].