AUTOMATIC DJ MIX GENERATION USING HIGHLIGHT DETECTION

Adrian Kim, Soram Park, Jangyeon Park, Jung-Woo Ha
Clova, NAVER Corp.
{adrian.kim, soram.park, jangyeon.park, jungwoo.ha}
@navercorp.com

Taegyun Kwon, Juhan Nam
Graduate School of Culture Technology, KAIST
{ilcobo2, juhannam}
@kaist.ac.kr

ABSTRACT

We demonstrate a fully automated system that generates DJ mixes from an arbitrary pool of audio tracks. Our goal is to offer a framework which automatically provides a high-quality endless mix of musical tracks without user supervision, given a seed track. To achieve this, we follow a filter and rank-based approach which selects the best-matching clip of a song to mix with the given seed track. We applied state-of-the-art musical information retrieval (MIR) methods including deep learning-based approaches such as automatic highlight extraction and key-tempo estimation for extracting cue points, choosing transition methods, and selecting audio clips. The results show that the proposed system can automatically generate competitive DJ mix contents plausible for users to enjoy.

1. INTRODUCTION

Continuously playing series of music clips as a single track, also referred as DJ mixing, is very common and popular in places such as clubs or electronic digital music (EDM) festival events. Making a DJ mix is an artistic job which requires high level techniques and knowledge. Thus, mixing has been mostly performed by professional disk jockeys (DJs). However, through recent advancement of music information retrieval (MIR) techniques, many DJ mixing tools now provide many features which lower the task difficulty for many users who are interested in musical creativity.

Although these features enabled the users to reduce the barrier of mixing, they still need to make their own mix by choosing music clips, selecting transition methods manually. Given a seed track which a user wants to start with, we automatize the entire mixing process of making an endless DJ mix, so one can create and enjoy a high-quality experience without making effort from selection tasks and using prior knowledge.

There exist some previous studies which aim to construct automatic mixing systems, but have several disadvantages. Cliff [3] applied beat-matching and time-stretching to mixing, but did not handle song selection and clip segmentation. Davies et al. [4] designed a mixing system using multiple features, but only remixes a single song by overlapping segments selected based on per-beat similarity computed on the entire song pool.

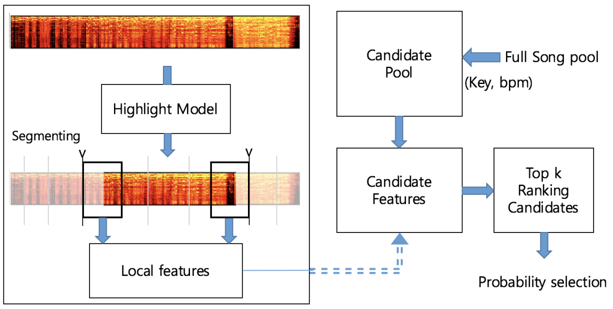

Figure 1. System overview for next clip selection

The contribution of the proposed system can be summarized, compared to previous methods:

- Given a seed track, the next music clip can be selected by a fast and simple end-to-end process.
- Many features used to select the next track are extracted from state-of-the-art methods based on deep neural networks.
- Clip selection is performed by utilizing a highlight detection model so that the mix plays perceptually interesting music clips [6].

2. SYSTEM DESCRIPTION

The proposed framework consists of the following four key components as shown in Figure 1.

2.1 Candidate pool selection

In order to reduce computation costs, given a seed track s, we first filter tracks to make a candidate subset C of the whole song pool T using simple rules with key and tempo values.

Harmonic mixing was done based on the Camelot Wheel [1] using estimated key values. Tempo matching was

Attribution: Adrian Kim, Taegyun Kwon, Soram Park, Jangyeon Park, Juhan Nam, Jung-Woo Ha. "Automatic DJ mix generation using highlight detection", Extended abstracts for the Late-Breaking Demo Session of the 18th International Society for Music Information Retrieval Conference, Suzhou, China, 2017.

[1] The Camelot Wheel was made by Mark Davis(http://www.harmonic-mixing.com/HowTo.aspx)

done by filtering tracks by bpm values with a small window size of w. Key and tempo values for all tracks were estimated using state-of-art methods [1] [5].

$$s \in T, C = K \cap B$$

$$K = \{t \mid key(t) \in CW(key(s))\}$$

$$B = \{t \mid bpm(t) \in [bpm(s) - w, bpm(s) + w]\}$$

2.2 Segmentation

Natural transitions between two music clips in a DJ mix is crucial to the user's experience. It is a non-trivial task to find specific points within a track that are used for transitions, which are called cue points. We use a beat-synchronized self-similarity based segmentation method inspired from [4] along with the downbeat tracking method from [2] to find these cue points, and produced segments from every track. Using highlight candidates extracted from networks made by [6], we select cue points that cover the highlights in order to make small, but significant candidate clips of the length of about one minute to play.

In detail, cue points are selected by finding peak values from a novelty function as in the following equation, where S is a downbeat synchronized spectrogram and $Ch(k)$ is a 2D-gaussian checkerboard kernel of size k.

$$Novelty[beat] = (Ch(k) * SelfSimilarity(S))_{beat,beat}$$

$$cuePoints = \{findPeak(Novelty)\}$$

2.3 Clip selection

Among the candidate segments, we use local features in order to get music clips that are most fit for the next song based on the given seed track. Local features include clip representations from a convolutional neural network trained on genre matching, energy values which are computed as mean mel-energy, loudness, spectral centroid features, etc. Given a ranking and mixability score based on local features, the final selection is done by picking the top-k candidates $C_k \subset C$ and sampling the next clip from a probability distribution P_t made proportional to the score.

$$P_t \propto Score(t), t \in C_k$$

2.4 Transition

The transition between the seed track and the selected clip is computed as a linear crossfade with the length l of four beats based on the average bpm of the connecting songs.

$$l = 4 * 60/\bar{bpm}$$

The process performed with the components above result in a mix of two music clips, and can be extended to make endless mix using the selected clip as a seed track. Performance can be a bottleneck if everything is done in real-time, so one can preprocess many of the features in order to enhance performance of the system.

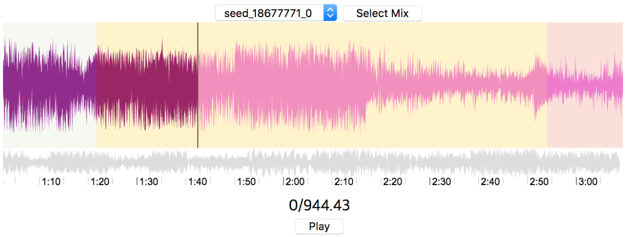

Figure 2. Example mix visualization from demo

3. DEMONSTRATION

Our demo presents the mixing contents automatically generated from the proposed system. Mixes were automatically generated on more than 1000 tracks from electronic music genres such as Trance, Techno and House added with some K-pop in order to see how well other genres mix together. Each clip used in the mix is visually separated with colors in order to represent transition points. Resulting audio mixes are pre-rendered for demonstration purposes.

4. FUTURE WORK

We are looking forward to evaluation metrics on mix quality, adapting more advanced mixing methods, real-time user interaction, and more machine learning based strategies in order to make a more intelligent system.

5. REFERENCES

[1] Sebastian Böck, Florian Krebs, and Gerhard Widmer. Accurate tempo estimation based on recurrent neural networks and resonating comb filters. In *ISMIR*, pages 625–631, 2015.

[2] Sebastian Böck, Florian Krebs, and Gerhard Widmer. Joint beat and downbeat tracking with recurrent neural networks. In *ISMIR*, pages 255–261, 2016.

[3] Dave Cliff. Hang the DJ: Automatic sequencing and seamless mixing of dance-music tracks. *Hp Laboratories Technical Report Hpl*, 104:1–11, 2000.

[4] Matthew E.P. Davies, Philippe Hamel, Kazuyoshi Yoshii, and Masataka Goto. AutoMashUpper: Automatic creation of multi-song music mashups. *IEEE/ACM Transactions on Speech and Language Processing*, 22(12):1726–1737, 2014.

[5] Ángel Faraldo, Sergi Jordà, and Perfecto Herrera. A multi-profile method for key estimation in edm. In *AES International Conference on Semantic Audio*, Erlangen, Germany, 22/06/2017 In Press.

[6] Ha Jung-Woo, Kim Adrian, Kim Dongwon, Kim Chanju, and Park Jangyeon. Music Highlight Extraction via Convolutional Recurrent Attention Networks. In *Proceedings of the 34th International Conference on Machine Learning, Machine Learning for Music Discovery Workshop*, 2017.