M-DJCUE: A MANUALLY ANNOTATED DATASET OF CUE POINTS

Mickaël Zehren
RWTH Aachen University
zehren@aices.rwth-aachen.de

Marco Alunno
Universidad EAFIT Medellín
malunno@eafit.edu.co

Paolo Bientinesi
Umeå Universitet
pauldj@cs.umu.se

EXTENDED ABSTRACT

A DJ-mix is an uninterrupted sequence of music tracks, constructed by transitioning from one track to another. The transitions are generally obtained by juxtaposing and overlapping two (or more) tracks. Needless to say, the locations in a track when the transition can take place have to be chosen carefully. These locations are normally referred to as "cue points", and their selection is one of the tasks that DJs perform when preparing a show. In particular, it is known that DJs use two objective functions to decide where transitions may occur: I) what we refer to as "extra-mixability" [1, 3, 4], which computes a compatibility score between two tracks (or segments of tracks), and II) the "intra-mixability" [1, 2, 5, 7, 8], which tells how well suited a specific position is for transitions, independently of the next track in the DJ-mix.

Our research focuses on the development of algorithms for the identification of cue points; for testing purposes, we built a dataset to be used as ground truth. To do so, we involved several DJs and musicians, to annotate tracks with cue points. For this dataset, only the intra-mixability is take into account as this criterios is more general, and it makes it possible to create annotations on a track-by-track basis.

While previous works on cue point identification exist, we feel that a convincing approach to evaluate different algorithms is still largely missing. Of those who addressed the problem, [2, 5] did not evaluate the quality of the cue points identified; in [1], the quality of cue points is assessed by evaluating 48 transitions, but neither the transitions nor the cue points used were disclosed; the evaluation in [8] is limited to checking whether or not the cue points detected are aligned with the 8-bar-period music structure; finally, although the approach in [7] is objectively evaluated, the dataset (30 tracks annotated by audio branding professionals) is not publicly available.

As part of our research, we are curating a dataset that currently contains annotations for 135 tracks of Electronic Dance Music (EDM), selected from a period of 30 years (1987–2016) and a variety of musical subgenres; about 60% of the tracks come from the digitalization of vinyls, as this is the format of part of the collection our dataset comes from. We believe that this mixture of audio sources is desirable to cover a wide range of situations.

It must be acknowledged that the locations of cue points in a track are subjective and ambiguous; indeed, they depend both on the taste and style of the DJ, and on the peculiarities of the track itself. To mitigate the impact of subjectivity, thus improving the replicability of the annotations, we took two measures: First, annotators were given a list of guidelines to follow when annotating (Table 1). Second, for each annotation, based on the agreement among the annotators, a confidence score was computed as the ratio of the people annotating this position and the number of annotators for the track. This score can then be used to remove annotations that did not find a high agreement between annotators. To make the confidence score meaningful, the dataset was independently annotated by five different musicians with a level of qualification ranging from computer-music scientist to semi-professional DJ, and we paid attention that each track was annotated by at least two people. Due to the large number of points for possible annotations, we decided to constrain cue-in annotations to the intro of the song only (see Guideline 4 in Table 1). In the same manner, the cue-out annotations are excluded from the intro of the track (see Guideline 5). This enables us to limit the time of the annotation process while keeping what we believe are the most valuable and commonly used cue points.

1. Mark the exact position in the track when you would switch tracks during a DJ-mix: This means that either this track or an upcoming track become prevalent in a cross-fade. The former is a cue-in point, and the latter is a cue-out.
2. Annotations coincide with positions of high novelty that do not interrupt strong musical elements (e.g., a bass or voice line). They are located at the start of a bar and are precise at 1/10th of a second.
3. Cue-ins are found where a track is considered to be able to stand on its own in the mix (when the musical content following this position is judged interesting and not too quiet to be played alone).
4. Cue-ins are found within the intro section of the track up to, but not exceeding the first beat of the core section (also called the main part or chorus).
5. Cue-outs are found anywhere within the track, but not before the core section.
6. Cue-outs can have a duration if a fade-out lasting up to the next major events in the outgoing track follows (e.g., a cue-out during a breakdown which precedes a chorus).

Table 1: Guidelines given to the annotators.

In our opinion, the usefulness of this dataset resides in the quality of the annotations and in the number of tracks annotated. This comes as a trade-off with the likely lack of coverage of valid candidate points resulting from the confidence filtering and the limited search space of the annotations. In other words, the set of annotations is not meant to be complete. We argue that this limitation could, and should be addressed when using this dataset for an evaluation by artificially limiting the assessment to the portions of the track covered by the annotations.

The dataset is freely available[1] following the JAMS specification from [6]. Because the tracks annotated are copyrighted, the audio content cannot be distributed. However, to make this dataset easy to be used, a tool is developed to automatically align the annotations to the new audio by detecting fingerprints in the tracks.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the Deutsche Forschungsgemeinschaft (German Research Association) through grant GSC 111, and from the German Academic Exchange Service (DAAD).

REFERENCES

[1] Rachel M Bittner, Minwei Gu, Gandalf Hernandez, Eric J Humphrey, Tristan Jehan, P Hunter McCurry, and Nicola Montecchio. Automatic Playlist Sequencing and Transitions. In *ISMIR*, pages 472–478, 2017.

[2] Dave Cliff. Hang the DJ: Automatic sequencing and seamless mixing of dance-music tracks. Tech. report, HP Laboratories, 2000.

[3] Roman Gebhardt, Matthew Davies, and Bernhard Seeber. Psychoacoustic Approaches for Harmonic Music Mixing. *Applied Sciences*, 2016.

[4] Tatsunori Hirai, Hironori Doi, and Shigeo Morishima. MusicMixer: Automatic DJ System Considering Beat and Latent Topic Similarity. In *MultiMedia Modeling*, volume 9516, pages 698–709. 2016.

[5] Adrian Kim, Soram Park, Jangyeon Park, Jung-Woo Ha, Taegyun Kwon, and Juhan Nam. Automatic DJ mix generation using highlight detection. In *ISMIR, late-breaking demo paper*, 2017.

[6] Brian McFee, Eric J Humphrey, Oriol Nieto, Justin Salamon, Rachel Bittner, Jon Forsyth, and Juan P Bello. Pump up the JAMS: v0.2 and beyond. In *ISMIR*, 2015.

[7] Diemo Schwarz, Daniel Schindler, and Severino Spadavecchia. A Heuristic Algorithm for DJ Cue Point Estimation. In *Sound and Music Computing (SMC)*, Limassol, Cyprus, 2018.

[8] Len Vande Veire. From raw audio to a seamless mix: an Artificial Intelligence approach to creating an automated DJ system. Master's thesis, 2017.

[1] https://github.com/MZehren/M-DJCUE