Unsupervised Detection of Structural Changes in Electronic Dance Music
Bruno Rocha1, Niels Bogaards2, Aline Honingh1

B.M.Machado@vu.nl, niels@elephantcandy.com, A.K.Honingh@vu.nl

We present a new system for the unsupervised detection of structural changes in Electronic Dance Music (EDM) based on timbre. We show how a musically informed approach can improve the accuracy of the algorithm. Results are in line with state-of-the-art methods that perform the same task.

Introduction
Music structure segmentation is a relevant topic in music information retrieval as it is both useful to the analysis of musical structure and a means to improve performance in other tasks [1], such as audio editing in recording workflows [2].

Recent years have seen Electronic Dance Music (EDM) gaining access to mainstream charts. However, it is still a neglected genre in music research with just a few articles published in journals and conferences. One exception is Butler, who claims that EDM presents a distinctive overall formal structure [3] and is not harmonically driven [4].

1. Detection of first downbeat

![Image](image1.png)

At this stage, we:
- load and downsample the audio;
- band-pass filter the signal in the low-frequency region;
- compute the energy on consecutive 30-second windows in order to determine the window in which the first downbeat is most likely to occur; we assume a salient kick drum is present at this moment;
- perform an onset detection on the selected 30-second window and identify the first downbeat.

This is a critical step, as we only need the remainder of the track (from the first downbeat onwards) for subsequent steps of the algorithm.

2. Tempo estimation and confidence measure

We combine two strategies to estimate the tempo:
- autocorrelation function (AC) of the onset detection curve;
- spectral decomposition of the same curve. The AC is translated into the frequency domain in order to be compared to this spectrum curve, and the two curves are subsequently multiplied [7].

Finally, a peak picking is applied to obtain a binary confidence measure is determined. This value is based on the harmonic relation between the detected peaks.

The confidence measure determines whether the musically informed rules (section 4) will be applied. For the tested datasets, more than 90% of the tracks had a confident tempo estimation.

3. Identification of structural changes

We follow Foote’s approach [5] with some modifications:
- magnitude spectrum is computed for each frame using a Fast Fourier Transform (FFT); the frames are beat-aligned with 83.7% overlap;
- contrasts to most common approaches, we perform a capstagram analysis in order to find periodic sequences in the signal;
- we then compute the cosine distance between each possible pair of frames in the cepstrum of data to get a self-similarity measure;
- consulting along the main diagonal of the similarity matrix using a Gaussian checkerboard kernel yields a novelty curve indicating the temporal locations of significant tonal changes;
- positions of prominent peaks of the novelty curve are selected as candidates for segment boundaries.

The result of this step is a segmented track.

4. Musically informed rules

Finally, for the tracks of which the tempo was estimated with confidence, we propose a set of heuristic rules to align the obtained novelty peaks with the beats. We analyse the distances between peaks and update them at each iteration, such as to form a dynamic structure.

Evaluation

![Table](image2.png)

Table 1 shows the results obtained for different datasets using the same parameter settings. It performs well on an in-house EDM dataset and shows how musically informed rules can improve the accuracy of the algorithm.

Although this method was created specifically for EDM, results on the RWC Pop dataset [8] can compete with the best performing algorithm submitted to WNEK 2012. This suggests that structural changes in pop music might have the same periodicity as in EDM. This method does not reach high performance on the Eurovision dataset. An explanation for this might be that pop music is mixed with traditional music from European countries in this song contest.

Conclusions

To conclude, we have presented a novel system that detects structural changes based on timbre. The success of the method is in agreement with Butler’s suggestion that EDM rejects harmony as primary musical parameter [3]. We have also shown how a musically informed approach can improve the accuracy of the algorithm.

Future Work

This system is part of a larger project about music similarity. This project, situated at the Centre for Digital Humanities, is a collaboration between the University of Amsterdam and Elephantcandy, a company that designs mobile audio and music applications.

![Diagram](image3.png)

Figure 10: Diagram of the project methodology