# TOWARDS EXTRACTION OF GROUND TRUTH DATA FROM DJ MIXES

**Diemo Schwarz**

UMR 9912 STMS IRCAM–CNRS–UPMC

**Dominique Fourer**

UMR 9912 STMS IRCAM–CNRS–UPMC

## ABSTRACT

DJ techniques are an important part of popular music culture but that are so far not very well researched because of the lack of annotated databases of DJ mixes. We first offer an overview of the necessary components to annotate recorded mixes automatically, which are: fingerprinting to obtain the tracklist, alignment to determine where in the mix each track starts and what tempo changes are applied to acheive beat-synchronicity, unmixing to estimate the fade curves for volume, bass and treble, and content and metadata analysis to derive the genre and social tags attached to the music to inform about the choices a DJ makes when creating a mix. Most of these components have been addressed by recent MIR research, except the alignment part for which we will give a first attempt using multi-scale correlation and dynamic time warping.

## 1. INTRODUCTION

This submission offers one missing brick in a larger research agenda to understand DJ practices—an important part of popular music culture. The outcomes from such an understanding are many, for instance musicological research in popular music, cultural studies on DJ practice and reception, music technology for computer support of DJing, automation of DJ mixing for entertainment or commercial purposes. So far, DJ techniques are not very well researched for the lack of annotated databases of DJ mixes.

In order to be able to annotate recorded mixes automatically, several components are needed, see figure 1:

**Identification** of the contained tracks (e.g. fingerprinting) to obtain the playlist,

**Alignment** to determine where in the mix each track starts and stops

**Time-scaling** to determine what tempo changes were applied to achieve beat-synchronicity,

**Unmixing** to estimate the fade curves for volume, bass and treble, and the parameters of other effects (compression, echo, etc.)

**Content and metadata analysis** to derive the genre and social tags attached to the music to inform about the choices a DJ makes when creating a mix.
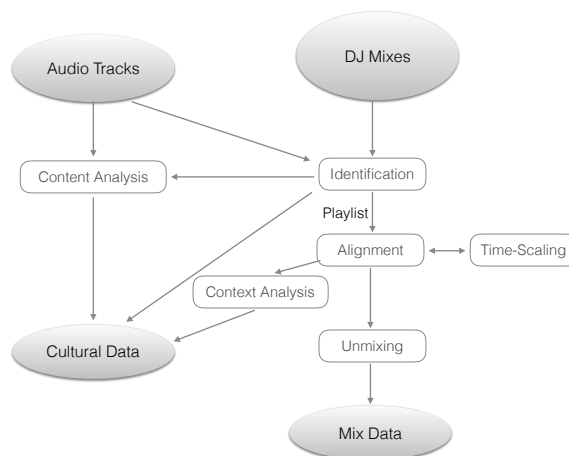
**Figure 1**. Overview of the larger context of information retrieval from DJ practices.

Most of these components have been addressed by recent MIR research, except the alignment part for which we will give here a first attempt using multi-scale correlation and dynamic time warping.

With some refinements, a massive amount of training data extracted from the vast number of collections of existing mixes could be made amenable to research in DJ practices, cultural studies, and automatic mixing methods.

## 2. RELATED WORK

First of all, there is much more existing research in the field of *studio mixing*, where a stereo track is to be produced from individual multi-track recordings and software instruments by means of a mixing desk or DAW, e.g. [2]. This research field has produced ground truth databases [3] and has some overlap with DJ mixing, when we see the latter as mixing just two source tracks, but the studied parameters and influencing factors differ too much from what is needed for DJ mixing. There is quite some existing work on tools to help DJs produce mixes, e.g. [1, 6, 7], but much less regarding information retrieval from recorded mixes. The first work opening up research on information retrieval from DJ mixes [9] tackles the identification of the tracks within the mix by fingerprinting. The authors also produce a database of ground truth annotations of playlists with rough start and stop times of tracks. Ramona and Richard [8] tackle the unmixing problem for radio broadcast mixes, i.e. retrieving the fader positions of

the mixing desk for several known and one unknown input signals, having sample-aligned source signals at their disposal. There is rare work on the inversion of other processing applied to the signal [4], notably compression [5].

## 3. MIX ALIGNMENT

The starting point for our method is the result of the previous stage of identification and retrieval on existing DJ mixes (see figure 1), or specially contrived databases for the study of DJ practices: We assume a recorded DJ mix, a playlist (the list of tracks played in the correct order), and the audio files of the original tracks. Our method proceeds in two steps: We first perform a rough alignment of the concatenated tracks with the mix by DTW, and refine that alignment to close in to sample precision.

The rough alignment uses the MFCC data of the mix and the concatenated MFCCs of the tracks as input. The DTW alignment path not only gives us the relative positioning of the tracks in the mix, but also their possible speed up or slow down to achieve beat-synchronous mixing or smoother evolution of the tempo of the mix.

Given the rough alignment by DTW, we then search for the best sample alignment by shifting a window of the size of an MFCC frame, taken from the middle of the track, around its predicted rough frame position in the mix. The minimum sum of square distances then determines the sample alignment.

## 4. EVALUATION

We tried the method on 3 collections of DJ mixes. The first collection is artificial mix data generated by mixing short excerpts of songs with given start points and volume fades. Unsurprisingly, the system can retrive the start points precisely to the sample, since no phase or volume alterations of the digital signal took place.

The second collection has been provided by a project partner of the EU project ABC_DJ. [1] These are 2 lounge mixes of pop songs with mild tempo adaptations. The ground truth start/end points can be retrieved with an average error of 0.148 seconds, $\sigma = 0.36$.

The third dataset [2] has been constituted in [9]. Unfortunately, it does not give information about the start point of the track in the mix.

The claim for sample accurate alignment can still be verified by attempting to remove the aligned track from the mix: For this we need to subtract it and observe the resulting drop in energy.

Figure 2 shows this on a mix from [9]: we can observe that for all but the 2nd and last tracks the mix energy shows a drop of over 15 dB. Because of the missing sub-sample alignment, mainly the low-frequency material is suppressed.
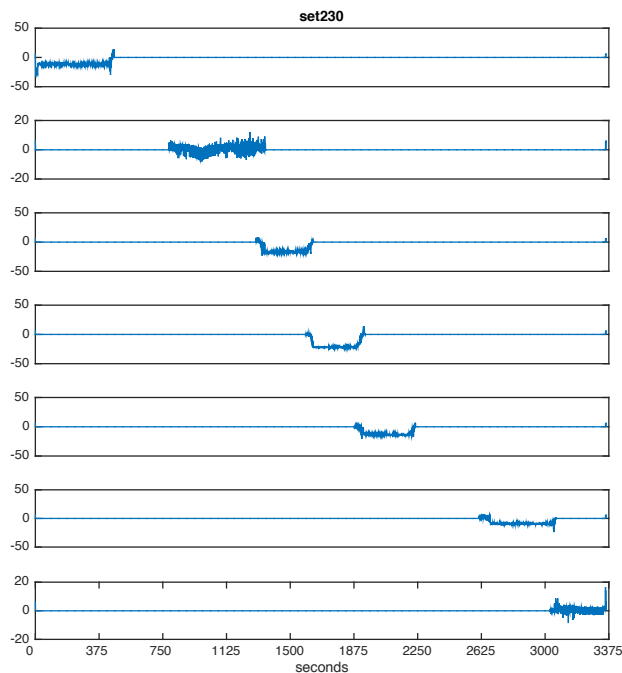
[2] http://www.cp.jku.at/datasets/fingerprinting



**Figure 2**. Suppression of tracks from a mix in dB.

## 5. REFERENCES

[1] Felipe Aspillaga, J. Cobb, and C-H Chuan. Mixme: A recommendation system for djs. In *ISMIR*, 2011.

[2] Brett Brecht De Man, R; King, and J. D. Reiss. An analysis and evaluation of audio features for multitrack music mixtures. In *ISMIR*, 2014.

[3] Brecht De Man, Mariano Mora-Mcginity, György Fazekas, and Joshua D Reiss. The open multitrack testbed. In *Audio Engineering Society Convention 137*. Audio Engineering Society, 2014.

[4] Stanislaw Gorlow and Sylvain Marchand. Reverse engineering stereo music recordings pursuing an informed two-stage approach. In *DAFx*, 2013.

[5] Stanislaw Gorlow and Joshua D Reiss. Model-based inversion of dynamic range compression. *IEEE TASLP*, 21(7), 2013.

[6] Hiromi Ishizaki, Keiichiro Hoashi, and Yasuhiro Takishima. Full-automatic dj mixing system with optimal tempo adjustment based on measurement function of user discomfort. In *ISMIR*, pages 135–140, 2009.

[7] Pablo Molina, Martín Haro, and Sergi Jordá. Beatjockey: A new tool for enhancing dj skills. In *NIME*, pages 288–291. Citeseer, 2011.

[8] Mathieu Ramona and Gaël Richard. A simple and efficient fader estimator for broadcast radio unmixing. In *Proc. DAFX '11*, pages 265–268, September 2011.

[9] Reinhard Sonnleitner, Andreas Arzt, and Gerhard Widmer. Landmark-based audio fingerprinting for DJ mix monitoring. In *ISMIR*, New York, NY, 2016.